11-30-2013

An Augmentative Gaze Directing Framework for Multi-Spectral Imagery

Libby Hsiao

An Augmentative Gaze Directing Framework for Multi-Spectral Imagery

by

Libby Hsiao

B.Eng.Mgt. McMaster University, 2000

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
College of Science

Nov 30, 2013

Signature of the Author _____

Accepted by _____
Coordinator, M.S. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
COLLEGE OF SCIENCE
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Libby Hsiao
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science
College of Science

Dr. Jeff B. Pelz, Thesis Advisor

Dr. David W. Messinger

Dr. Reynold J. Bailey

Date

THESIS RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE

Title of Thesis:

An Augmentative Gaze Directing Framework for Multi-Spectral Imagery

I, Libby Hsiao, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____

Date

An Augmentative Gaze Directing Framework for Multi-Spectral Imagery

by

Libby Hsiao

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Master of Science Degree
at the Rochester Institute of Technology
College of Science

Abstract

Modern digital imaging techniques have made the task of imaging more prolific than ever and the volume of images and data available through multi-spectral imaging methods for exploitation is exceeding that which can be solely processed by human beings. The researchers proposed and developed a novel eye movement contingent framework and display system through adaption of the demonstrated technique of subtle gaze direction by presenting modulations within the displayed image. The system sought to augment visual search task performance of aerial imagery by incorporating multi-spectral image processing algorithms to determine potential regions of interest within an image. The exploratory work conducted was to study the feasibility of visual gaze direction with the specific intent of extending this application to geospatial image analysis without need for overt cueing to areas of potential interest and thereby maintaining the benefits of an undirected and unbiased search by an observer.

Acknowledgements

I would like to thank Dr. Jeff B. Pelz for his guidance as a professor and advisor and his tutelage and guidance in conducting psychophysical studies. His feedback in the development of the gaze assist methodology was instrumental in advancing it beyond what had been initially envisioned, and the opportunity to pursue work in this new area of research with the Multidisciplinary Vision Research Lab.

Thank you to Dr. Reynold J. Bailey and his student Srinivas Sridharan for allowing the use of their previously developed subtle gaze direction framework and assistance in adapting the framework for use with mutli-spectral imagery.

Thank you to David W. Messinger and Amanda Ziemann for their assistance in helping me understand their work with the spectral complexity metric algorithm and encouraging me to actively pursue research with it.

Most importantly, thank you to Alison Durocher for her love, patience and support, without which this would not have been possible.

Contents

List of Figures

Chapter 1

Background

1.1 Introduction

Modern imaging systems and digital methods have made the task of producing images easier and more prolific. Not only has this trend been noticed in commercial and consumer applications, but by the scientific community at large. New real-time ground imaging technologies capable of gigapixel image capture have been deployed with an expected increase of 70,000 percent increase in data according to analysts' estimates (Jean, 2011). Similarly, Andriole, Wolfe, Khorasani, Treves, Getty, Jacobson and Seltzer (2011) emphasized in their review the sheer volume of the large data sets which can now result for a single patient's computer tomography (CT) scan presents unprecedented challenges in analyzing and visualizing despite the abundance of information not available previously that is gained. This underpins the argument and need for new and innovative ways for end users to navigate and visualize complex data sets.

In visiting the potential design of gaze assist methods for image presentation and target identification to geospatial analysts, a literature review of the automation in decision support systems, potential bias effects and relationship between human decision bias and

automated decision support systems relating to visual search task performance was carried out. While the straight forward approach to aiding an observer may be to highlight potential targets of interest, this may introduce highly salient areas to an image that distracts an analyst from performing a thorough search. There is also evidence of research that this can induce other forms of errors or confirmation bias from over-confidence in the underlying automation and/or complacency effects. Instead, the researchers proposed a system whereby an observer's gaze be cued to regions of interest with minimal awareness on the part of the observer and reduced awareness of assistive automation to study the gaze patterns as compared to an unaided control group.

1.2 Objectives

The objectives for this research project were to develop an eye movement contingent display and integrate this with image processing algorithms to act as the basis for a subtle cueing method to direct observer gaze to specific regions of an image. The main objectives of this thesis were to:

1. Implement an eye movement based display
2. Integrate spectral pre-processing algorithms as a way of determining regions to be cued within a multi-spectral image
3. Dynamically update regions to be cued based on real time observer gaze patterns
4. Investigate the efficacy of the subtle gaze direction in the context of remotely sensed images
5. Determine if subtle gaze cues yield improvements in visual search tasks within a remotely sensed images

1.3 Automation, Decision Support and Operator Bias

The introduction of automated decision support systems has lead to a broad area of research and active discussions on best practices of their employment. Questions with regard to the kinds of tasks and extent of automation to be implemented in a specific workflow or environment are some examples (Parasuraman & Riley, 1997). The work of Skitka, Mosier and Burdick (1999) provides evidence that there is improved performance and a reduction of errors in a simulated environment of monitoring and tracking tasks involved in flying commercial aircraft with automated monitoring aids. Failure to detect an event however, in the presence of an automated aid, led to increases in errors relative to a non-automated environment. Extensive research in the area of automation systems and bias effects that can impact operator decision has produced mounting evidence that an over reliance on automation methods can result in complacency and "automation bias", where two common classes of errors that emerge are:

1. Errors of omission: a failure to identify irregularities missed by the underlying automated decision support system.

2. Errors of commission: falsely identifying irregularities as suggested by the automated support system.

The two broad classes of errors could simply be taken as the false alarm and positive hit rates of the automated task without intervention. However, with a human in the loop, the information is presented as a decision aid and subject to an operator's decision to act on the information presented. Research conducted by Meyer (2004) has shown whether action is taken by an operator is dependent on the inherent trust in the underlying automation processes. Under this type of assumption, one can model the uncertainty in automation methods using signal detection theory (SDT) and Bayesian statistics. In his study to con-

ceptualize operator behavior with respect to an automated warning system, he introduced two separate responses to a binary warning system: compliance and reliance. Compliance was defined as the degree an operator would act in accordance with a warning cue present, while reliance referred to the degree to which an operator dared to not act in the absence of a warning cue.

In a review conducted by Wiczorek, Meyer and Guenzier (2012) which investigated the relationship between these two trust mechanisms with respect to an aided visual search task, a study was carried out by independently varying the statistical probability of there being an event present when the automated system has alarmed. The authors coined the terminology of Positive Predictive Value (PPV) and Negative Predictive Value (NPV), which can be expressed mathematically as of $p[Signal|Alarm]$ and, $p[noSignal|noAlarm]$ respectively, of which they varied the rates independently. The results demonstrated an asymmetry in the separate trust indicators and while the two mechanisms are distinctly separate, were shown to be somewhat related. Differences in NPV had no effect on operator compliance and a high PPV was the primary factor in eliciting a response. In contrast, evidence of system reliance was not present until PPV was high, indicating an operator's likelihood to trust that everything is fine takes into the consideration the accuracy of a system to detect trigger and alert of an event. The authors hypothesized this asymmetry in system trust was a result of operators placing greater cognitive saliency to false alarm events as supported by the previous work of Rice and McCarley (2011).

While this points to the importance of minimizing the false alarms in an automation decision aid in order to gain operator compliance, a force which acts in opposition to this is that of the satisfaction of search (SOS). Satisfaction of search is a phenomenon that is of particular interest under the broader class of error of omission and one that has long been a concern in the field of radiology (Rogers, 2000). Satisfaction of Search occurs when a

positive detection of a target can reduce the subsequent detection of additional targets in the same search instance that is not identified. While traditionally studied under the domain of medical imaging, the phenomenon has been extended to generalized SOS to include visual detection tasks to applications such as airport luggage screening (Fleck, Samei, & Mitroff, 2010) and in military command and control environments (Yeh, Wickens, & F, 1999). Therefore, not only does a faulty automation algorithm adversely affect an operator to reduced rates of user compliance with lower PPV values, but an algorithm with low false alarm rates and high PPV which fails to detect all possible targets has the potential to lead to SOS by the observer.

In summary, although thorough understanding of automation bias is a complex interaction subject to individual operator differences, acceptance of risk, and is specific to the nature of the task, there is little disagreement regarding its potential detriment when over-reliance and complacency results and the importance of minimizing the potential impact. This is especially the case when it comes to support systems where decisions being made can result in catastrophic loss to human life or to the environment. Not only is mitigating these effects through system design and training desirable, but should be a key consideration to the overall system design.

1.4 Spectral Imaging

Modern remote sensing platforms are capable of imaging in the electromagnetic spectrum beyond the wavelengths visible to the human eye (400-700nm). The image data collected can then be exploited to augment human capabilities for event or target detection. In multi or hyperspectral imaging electromagnetic radiation of a scene is captured and focused onto a spectral grating and disperses the energy of the specific wavelengths to a corresponding detector. The specific number of channels for a given system can number from several

broad spectrum bands to in excess of 200 densely sampled distinct wavelengths. The sensor records radiance which with calibration data can then be converted to establish the surface reflectance for each spectral band as the sensor passes overhead. This culminates in a n-band image, where n is the number of discrete sampling channels/wavelength bands.

The resulting data set typically requires spatial registration and then is considered as an image cube where each pixel site can be thought of as a spectrum of the sampled area (Fig 1.1). The spectrum at each pixel can then be represented as a n-dimensional vector and analyzed using various mathematical constructs.

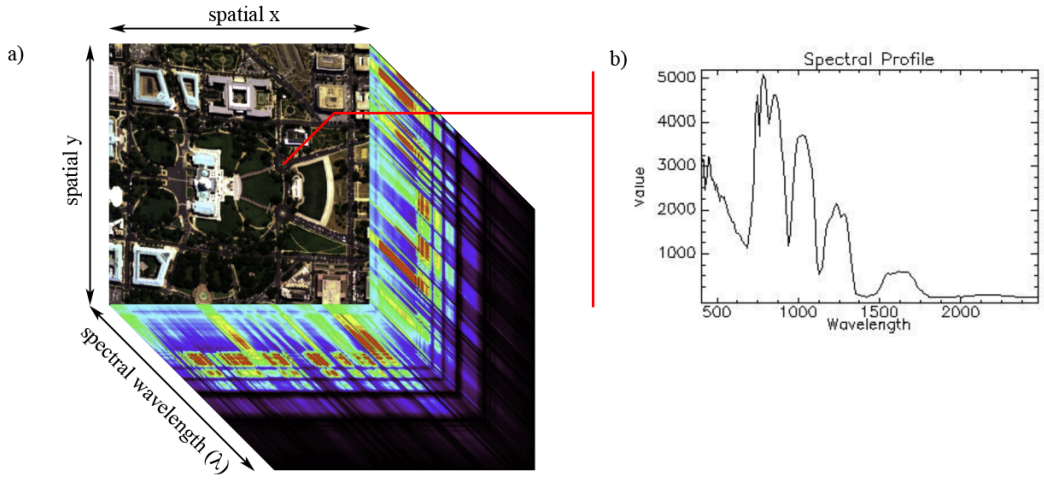


Figure 1.1: Graphical representation of: a) hyperspectral image cube b) recovered spectra at image pixel location

The design challenge that exists for a multi or hyperspectral imaging platform is to balance the tradeoff in performance characteristics in terms of spectral sampling density (number of spectral bands) and that of the spatial resolution. Since the radiance of the sampled area is divided spectrally the higher the number of sampling bands, the more of a light starved situation arises. While densely sampled spectral data can be exploited further with mathematical analysis, larger sensor pixels are typically required to capture

more photons. This in turn, corresponds to larger ground sampling distance (GSD), the area of coverage for a single pixel which results in a higher likelihood to contain a greater variety of material and mixed spectral response. The combination of spectral and spatial resolution at which an image is sampled then, defines the extent to which the data can be exploited.

1.5 Spectral Analysis

Target detection algorithms typically fall under one of three major approaches: geometric/deterministic, statistical or an examination of the spectral feature spaces (Schott, 2007). The deterministic approach assumes that the spectra at any pixel location within an image can be represented by end members which are defined by the geometric extremes of the n-dimensional space and the 'purest' spectra in the image. Thus, the set of vectors formed by the end-members can be thought of as a non-orthogonal set of basis vectors such that the spectra at each pixel location can be represented as some linear combination of them. The statistical approach assumes the image data can be described by clusters of material classes when the spectra are plotted in a vector space and characterized by statistical measures such as the mean vector and covariance matrix. Finally, spectral feature detection seeks to look for specific absorption features of a known target or compare the pixel spectra with those of know targets of interest and is analogous to classical spectroscopy.

The data products produced by spectral analysis are generally class segmented images, a probability map at the pixel level of statistical anomalies or as compared to a known spectral signature. This is in contrast with image interpretation techniques as performed by humans which do not process images per pixel but rather relies on attending to portions of an image and makes use of the distinct visual cues of: shape, size, tone, texture, pattern, shadow and site as described by Estes and Thorley (1983). It is therefore not uncommon for

the data products to be graphically overlaid or cross-referenced to the visualized image data. However, the raw data that can now be gathered far exceeds what our computer display or human visual system can process. This being the case, significant dimensionality reduction and image processing and/or enhancement decisions are made by an image analyst. These techniques are not without trade-offs and analysts must be trained to be aware of the implications of their decisions in the visualized output for their interpretation.

These distinct differences present both a challenge and opportunity to leverage the strengths and flexibility to visual image interpretation with data products derived from the augmented capabilities of spectral imaging.

1.6 The Human Visual System

The experience of vision is a sophisticated system of sensing light from the immediate environment and visual processing at the cognitive level. One senses the environment through light passed through the physical optics of the eyes which then undergoes cognitive processing as the neural signal travels to the visual cortex. At a physical level the human eyes are controlled by six sets of ocular muscles that allow us to move the eyes to sample our surroundings without having to move the entire head. Once the photons have passed through the geometric optics, an image is formed on the retina, a curved surface at the back of the eye lined with two classes of photoreceptors: rods and cones. The rods are spread primarily throughout peripheral region reaching peak densities at approximately 20 degrees off-axis from the fovea, a pitted region in the retina (Fig 1.2). The rods are highly sensitive to light and responsible for vision during low illumination conditions while cones are responsible for all our experiences of color and high acuity under well lit conditions. The distribution of cones is highly concentrated at the center of the retina with a peak occuring at the fovea, where there are only cones and accounts for approximately a 1-2 degrees of the

visual field. This densely sampled region of the image on the retina allows for a resolution that is almost 100 times than at the peripheral field of view (Rao, Zelinsky, Hayhoe, & Ballard, 1997).

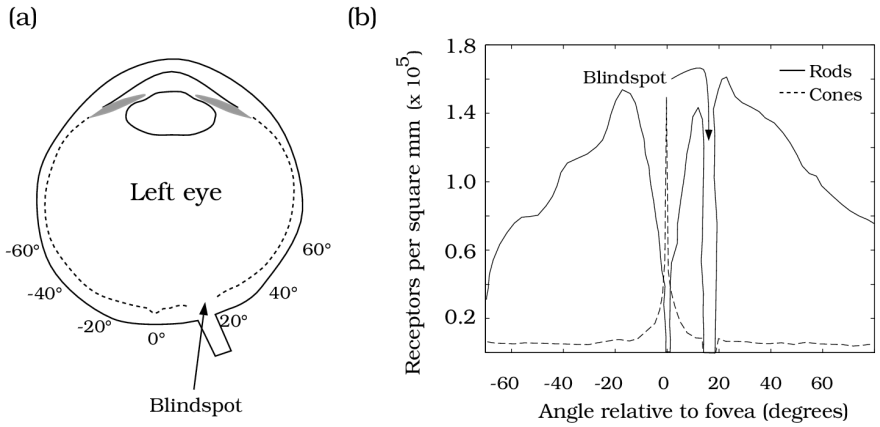


Figure 1.2: a) Visual Angle relative to the position of the Fovea (Left Eye). b) Distribution of Rods and Cones on the Retina of the Human Eye

In order to experience our environment with high resolution with our spatially resolution limited system, spatio-temporal integration must take place in the visual processing stage and our eyes must sample the scene rapidly. This rapid sampling is possible because the ocular muscles that control the eye are capable of moving it at speeds of up to 700 degrees per second (Rao, Zelinsky, Hayhoe, & Ballard, 1996). These ocular muscles initiate movements to stabilize the image on the retina, pursue a moving object, or redirect attention to a new region of the scene. In the process of exploring a visual scene, a person makes about three eye movements per second amounting to more than 160,000 a day (Schiller, Kendall, Slocum, & Tehovnik, 2008). These rapid movements known as saccades are pre-programmed ballistic movements that direct the eye to regions of interest that require high acuity and take 150-200 ms to plan and execute. The actual movement of the eye is performed in 20 ms + 2 ms per degree of visual angle during which visual perception is suppressed so that blurring of the scene is not perceived. In the absence of a saccadic movement, the eye is

in the process of a fixation and engaged in the process of acquiring information about the scene.

These eye movements are manifestations of selective visual attention and studying these eye movements allows one to understand how the visual attention is being deployed under various conditions and tasks. (Pelz & Canosa, 2001)

Chapter 2

Gaze Directing Framework for Multi-Spectral Imagery

2.1 Overview

The inherent risk of observer bias, the opposing interactions of compliance based trust arising from an automated algorithm and the phenomenon of satisfaction of search outlined in the review of automation bias, motivated the researcher to propose a flexible and scalable subtle gaze directing framework to assist image analysts. The intended objective was to allow an image analyst to search allowing study of images as they would without intervention while being assisted in their search with non-overt but directed cues to regions of interest. The regions of interest were established using multi-spectral image processing algorithms with the underlying source image data which may not be easily visualized. This framework can be conceptually described by four key modules working in conjunction with a human observer in the loop (Fig 2.1): an image preprocessing module to determine cueing coordinates, a logic module responsible for when and where a cue is initiated or suppressed, an

image rendering module to control the appearance of the cue and stimulus image and an eye tracking system to incorporate real time information of an observer to feedback to the system.

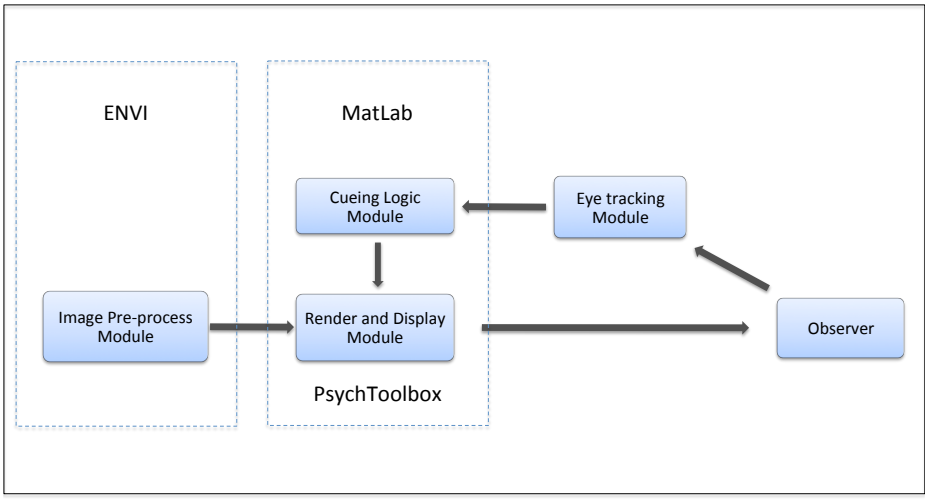


Figure 2.1: Framework Block Diagram

For the specific research goal of assisting geospatial image analysts, the researcher sought a preprocessing algorithm to be one that could mimic regions where experts would be inclined to put under scrutiny while the cueing method was carried out with minimal interference or awareness on the part of the observer. The description of these modules remains as generic functional blocks to emphasize that this framework could be scaled or evolved as improved target detection algorithms or improved cueing methods become available through continued research. Similarly, the cue control module which sets the conditions for when and the order with which cues are presented could easily be modified and adapted to the needs of specific applications. The researcher elaborated in further detail the specific methodologies and criteria for the logic that was employed in the proposed experimental system in the sections which follow.

2.2 Simulating Expert Gaze

Recently efforts have been made in studies to observe whether the performance of novices can be elevated by replicating the attentional gaze of an expert, leading to improved training efficiencies (Sridharan, Bailey, McNamara, & Grimm, 2011; Vine, Masters, McGrath, Bright, & Wilson, 2012). While studying gaze patterns for repeatable and relatively consistent tasks can yield patterns in expert gaze that can be leveraged, it is much more challenging and problematic in the case of image analysts. Unlike medical experts where the expert observer may be looking at numerous images of a general anatomic region sharing similar features and topologies, where optimized scan patterns may form; image analysts have the potential to be faced with varied and changing geographic content, terrain, and weather effects within the images themselves as well as the target of their search. Complicating matters further is information available for exploitation depending on the imaging sensor capabilities and decisions which can be made by an analyst as to how they choose to visualize the data and optimize it for a given task. The researchers hypothesized this makes an expert gaze highly variable based on a number of factors such as image terrain, the target of interest, and method of visualization, therefore making these gaze patterns difficult to generalize and simulate with an unsupervised target detection algorithm.

Recent work in metrics of spectral complexity with applications to large area search applications provides an interesting generalized approach to evaluate multispectral images and a basis to simulate expert gaze (Messinger, Ziemann, Basener, & Schlamm, 2012). In their work, the localized spectral 'complexity' is a measure defined by the volume of the convex hull required to contain the end-members found within a sub-tiled region of the image. This can be thought of as performing the unique task of evaluating image saliency based on spectral contrast of a given tile size relative to others, leveraging the augmented information available to the imaging sensor capabilities. Since this is not a specified target

search requiring fine spectral sampling methods, even an under sampled spectrum provides significantly more information than that available only through the visible spectrum which can be leveraged in combination with an analyst's visual inspection. Furthermore, unlike most target detection algorithms which evaluate target likelihood at the pixel level, the algorithm seeks to evaluate a region of the image itself. This algorithm showed promise and appeal to the research application as it shares similarities with the nature of analysis of regions of the image much like that of how human vision selectively attends to a foveal image that is experienced highest acuity.

The research proposed that computing the complexity map for a given image and thresholding a top percentile of the results had the potential to reveal generalized regions of interest to a human observer to confirm potential targets, independent of the nature of image.

2.3 Large Area Complexity Metric

The following section examines, how the volume of the convex hull is calculated and the trade-offs in the selection of variable parameters (i.e. number of end members and given tile size) in large area complexity measure is examined. In a physical sense, this mathematical construct works under the hypothesis that areas of greater material diversity are more likely to be of interest. The researcher began with an exploration of some characteristics of the Gramian matrix and its properties, where the Gramian of vectors $v_1, \ldots, v_k$ is defined as:

$$G(v_1, \ldots, v_k) = detM^*M \tag{2.1}$$

$$\text{where, } M = (v_1, \ldots, v_k)$$

The Gramian has the distinct property that the volume of the parallelotope formed by the end member vectors as its edges. In the original work of Messinger et al (2012), the effect of number of end members to iterate through was studied and demonstrated that the volume function decreased monotonically and reached zero once the end-member search began to return vectors that were not linearly independent (an inherent property of the Gramian). The volume function's peak, the value selected as the complexity metric, occurred at relatively lower dimensionality as a result of the Max-D end member algorithm which returned end members in decreasing magnitude. This suggested the computed complexity metric was unchanged, provided a reasonable number of end members were selected for the volume to be computed over.

2.3.1 Effect of Tile Size

The discussion which follows and the analysis of tile size on the complexity metric used sampled images consisting of a Quickbird 4-band image of a forested mountain scene in Esperanza, California, a pastoral scene of Rondonia, Brazil and a coastal scene off Salvador, Brazil the last two of which were acquired using Worldview-2 (8-band images) (Fig 2.2).

The effect of user selection of sub image tile size to evaluate has a complicated interaction between spatial resolution of the search, size of the tile cue and statistical separability for areas of interest. At the most fundamental level, the resolution of the imaging platform will limit the type of target search that is practical to conduct through visual analysis. For instance, trying to locate a single civilian in an image with 5m pixel resolution is not a practical task. At a minimum the target must be twice the size the spatial resolution of a pixel according to the Nyquist-Shannon sample theorem for alias free signal reconstruction. Since the large area search complexity metric is computed based on sub-image tiles of a specified size, the computed results for a given tile size are an indication of:

1. The variety of dissimilar material constituency changing within a given area and;

2. The ability to spatially localize potential targets of interest from the results.

In the case of the material composition within a given tile, if we consider the extreme case of a single pixel, only a single spectrum exists and the Gramian of a single vector is by definition its norm. In other words, it is a measure of the reflectance at the ground location of the pixel taking into consideration all sampled wavelengths and the resulting complexity map is a panchromatic image across the sampling space. While this provides maximal spatial resolution, it provides no information as to how regions of the image vary in complexity as demonstrated using a 4-band image from Quickbird scene of a controlled fire over Esperanza (Fig 2.3 (a)).

(a) Esperanza, California Scene

(b) Salvador, Brazil Scene

(c) Rondonia, Brazil Scene

Figure 2.2: Source Image Mosaics Shown Displayed Using Visible Bands at 2% Linear Stretch

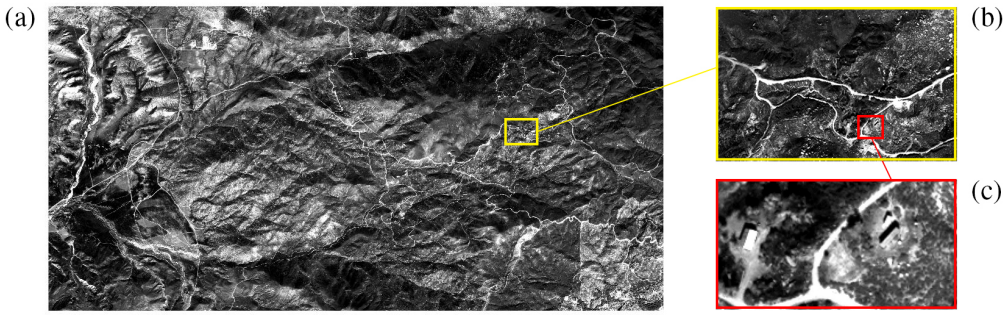
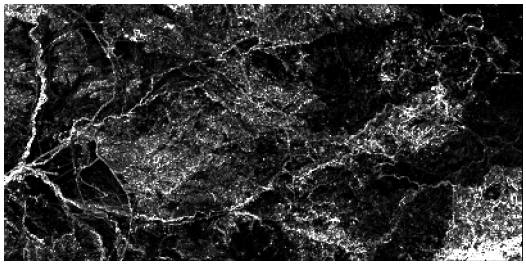


Figure 2.3: (a) Complexity map computed at 1 pixel tile (displayed at 2% linear stretch)
(b) Region of interest magnified from high resolution panchromatic image
(c) Building sheds visible under high magnification

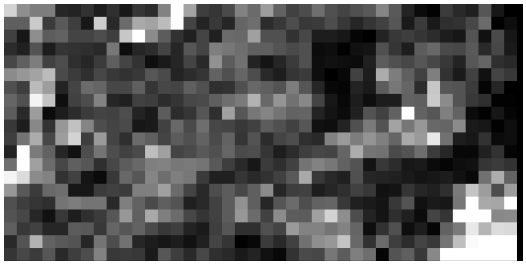
As the tile size is increased to 10 x10 pixels (Fig 2.4a), the image still maintains much of the feature resemblance to that of the one pixel tile case, but appears more pixelated and at a reduced spatial resolution. Thresholding the top five percent of the brightest pixels reveals the high concentrations of buildings at the top and bottom right of the Esperanza fire scene as well as the group of building sheds located toward the center of the right third of the image (Fig 2.3 (b) and (c)). As one continues to increase the size of the sub image tiles, the content of the image itself becomes less recognizable, but the result is an image where the brightness of the tile is a normalized brightness map of complexity. As the tile sizes are chosen to be increasingly large, potential targets of interest become more difficult to localize and fewer tiles are returned as part of the thresholding process (Fig. 2.4a-2.4f).

One can also seek to characterize algorithm behavior and performance as we change the tile size by examining the histograms of the complexity image. Since histograms are image content dependent, the task of fitting a specific curve of a functional form may not be feasible. Despite the plotted histogram of the complexity result bearing some resemblance to a normal distribution as tile sizes increase, since the histogram is bound by [0-255] by construct of the algorithm and is a measure of the volume of the convex hull containing the



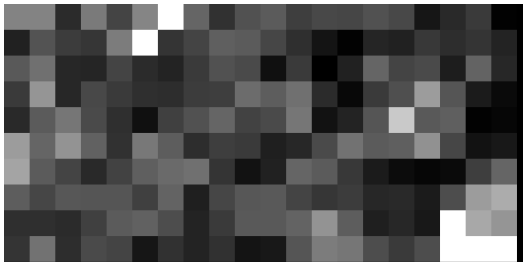
(a) 10 x 10 pixel tile complexity map

(b) top 5 percentile results of 10 x 10 pixel tiles



(c) 100 x 100 pixel tile complexity map

(d) top 5 percentile results of 100 x 100 pixel tiles



(e) 200 x 200 pixel tile complexity map

(f) top 5 percentile results of 200 x 200 pixel tiles

Figure 2.4: Complexity Map and Thresholded Results for Varying Tile Sizes

end members, and therefore cannot be negative valued by definition, a fit of the Gaussian form is not appropriate. The researcher submits instead that studying the behavior of the first four standardized moments of mean, standard deviation, skew and excess kurtosis defined by equations (2.3)-(2.5) respectively, may provide some insight as to how tile size impacts the computed complexity map. The four moments of interest are calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.2}$$

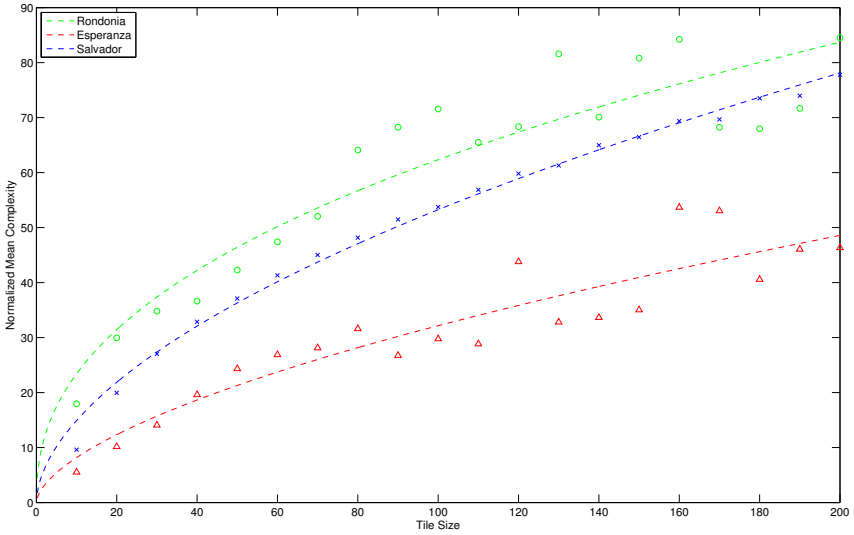
$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}} \tag{2.3}$$

$$skew(x) = \frac{E[(x - \bar{x})^3]}{\sigma^3} \tag{2.4}$$

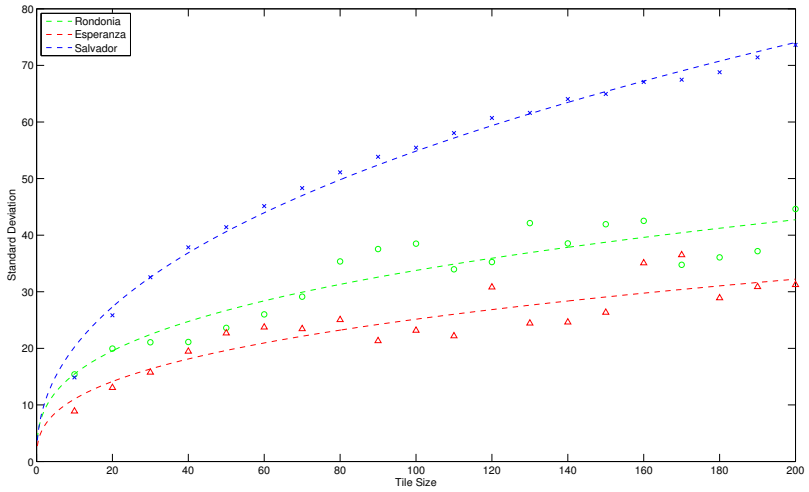
$$\text{excess kurtosis}(x) = \frac{E[(x - \bar{x})^4]}{\sigma^4} - 3 \tag{2.5}$$

The histogram metrics are computed for three large area calibrated test scene mosaics of different terrain types in 10 pixel intervals up to a tile size of 200 x 200 pixels. This gives a sense for the behavior of the overall metric as a function of tile size. The behavior of the four metrics with respect to increasing tile size was plotted in Fig 2.5 and Fig 2.8.

As tile size increased, the trend of the mean (Fig 2.5a) and standard deviation (Fig 2.5b) of the computed tile metric increased and plateaued similarly for all three images and took on the shape of a root function. This behavior was expected as the size of the tile was increased as one would expect that material composition found from tile to tile to be similar and therefore, tend toward a higher mean. However, once a sufficiently large area tile size was reached one would also expect the abundance and diversity of materials found



(a) Mean Complexity vs Tile Size



(b) Standard Deviation vs Tile Size

Figure 2.5: Standardized Moments 1 and 2 of Complexity Histogram

from tile to the next to be similar, and therefore the mean complexity metric to change less once a threshold tile size was reached. While this was the general trend, the root function modeled appeared to only be a good approximation for modest tile sizes (80 pixels or less). The researcher suspected this may be owing to the fact that as tile sizes increased it lead to potentially larger edge portions of the image which are not computed resulting in values that oscillate between the functional approximation.

It is useful to consider what the third and fourth standardized moments represent prior to discussion of their results. Skew is a measure of the extent of how far a probability distribution leans toward one side of the mean while excess kurtosis is an indicator of how peaked and long the tail of the distribution is. In general, for the case of a unimodal distribution, positive skew values tend to indicate an asymmetry of a larger area under the curve to the left and conversely, negative values to the right but in multi-modal distributions becomes more difficult to interpret (Fig 2.6).

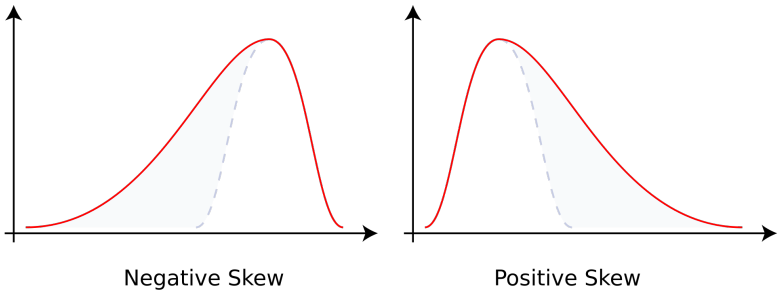


Figure 2.6: Positive and Negative Skew Shapes

Skew and excess kurtosis appeared to decrease exponentially, suggesting that the resulting complexity map histogram grew more symmetric about the mean value and that the high peak and heavy tail decreased rapidly with increasing tile sizes. The Rondonia pastoral scene was an exception to the case and plotting the histogram itself to glean an idea of how it was changing gave some explanation of why this may have happened. The histogram for small tiles sizes showed a second peak of relatively low complexity metrics

and as the tile sizes were increased the calculation of the skew metric increased due to the higher number of low complexity tiles despite the shape of the histogram becoming more symmetric and resembling a normal distribution (Fig. 2.7).

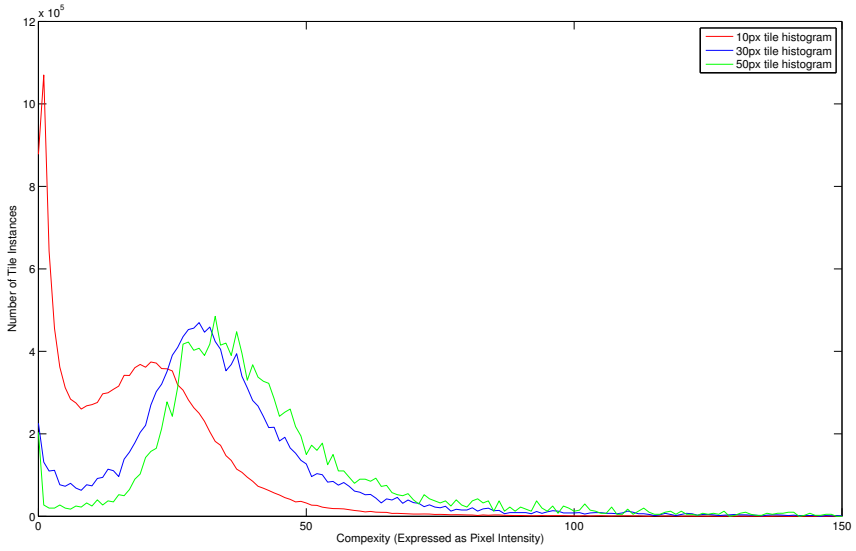
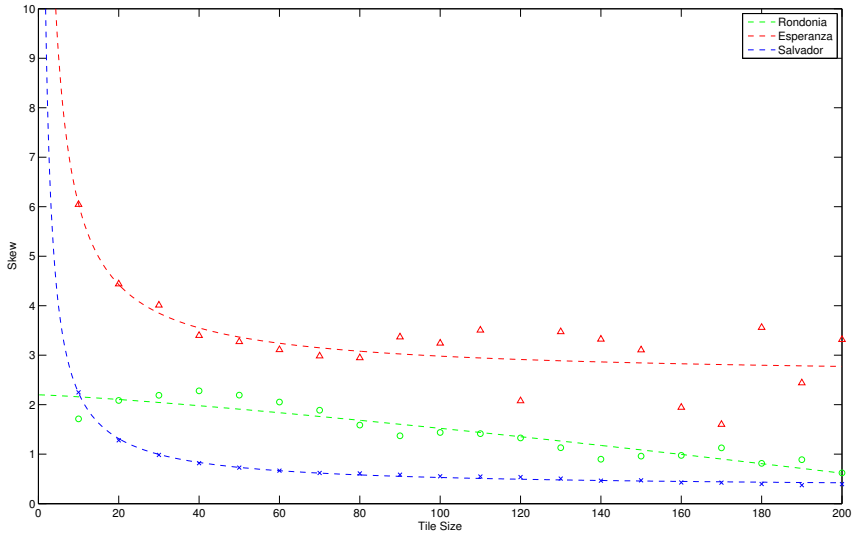
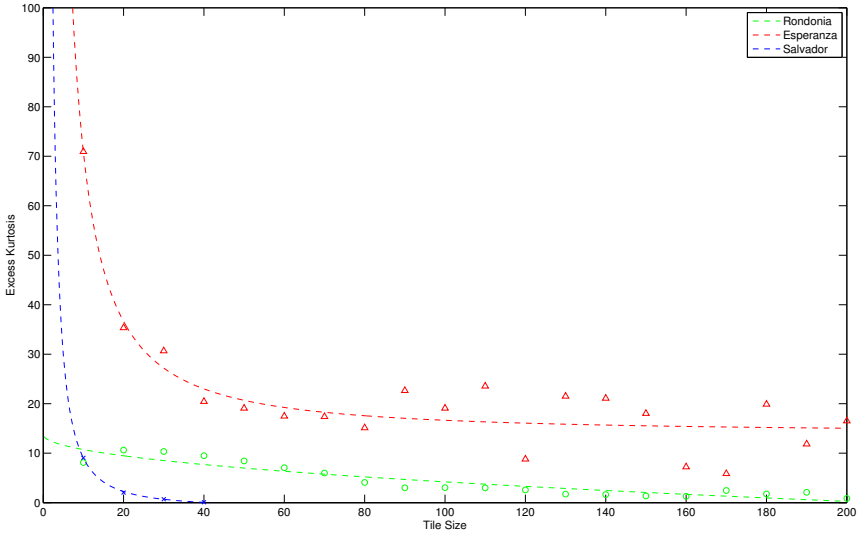


Figure 2.7: Histogram of Complexity Metrics for Rondonia Mosaic at 10, 30 and 50 pixels

Without a thorough analysis of a large sample of images to conduct a similar analysis on, it was not possible to state whether this was indicative of the unpredictable nature of the third and fourth standardized moments for images whose metric might have demonstrated two distinct peaks in the histogram. Although the decreasing skew and excess kurtosis suggested decreasing separability of higher complexity measure tiles from the mean of 'background' tiles and regions of interest by thresholding results, the tile sizes themselves are larger and fewer in the returned result. Therefore, the thresholding remains surprisingly robust as fewer tiles are returned, but of much larger regions. However this came at the cost of reduced spatial fidelity where the size of the tile may require a secondary localized search for the feature(s) of interest.



(a) Complexity Skew vs Tile Size



(b) Excess Kurtosis vs Tile Size

Figure 2.8: Standardized Moments 3 and 4 of Complexity Histogram

The fairly robust nature of the algorithm in identifying potential areas of interest suggested that the choice of tile size would be dictated by a trade-off between the specificity of localizing the targets within the smaller sub-image tiles against that of a more manageable number of results returned based on larger tile sizes. For the purpose of employing this to leverage the source image data to direct visual gaze in a search task, the size of the foveal image the human visual system can provide some guidance as to selection of tile sizing for computations. If one considers the case where gaze is directed to an image tile much larger than that of a corresponding foveal image, the feature(s) of interest which may occur within an area larger than the 2 degree visual angle of the fovea and result in an observer not attending to the feature in the region of the highest visual acuity. Examination of the feature of interest would then be subject to the condition that the observer must then initiate a localized search in the area. While this is a reasonable assumption given the nature of the task, the researcher submits that it would be more effective to direct an observer as close as possible and that under ideal conditions, the size of the sub-image tile chosen based on the display resolution and to match that of a 1-2 degrees of visual angle corresponding to that of the fovea.

2.3.2 Task Influence Based Complexity

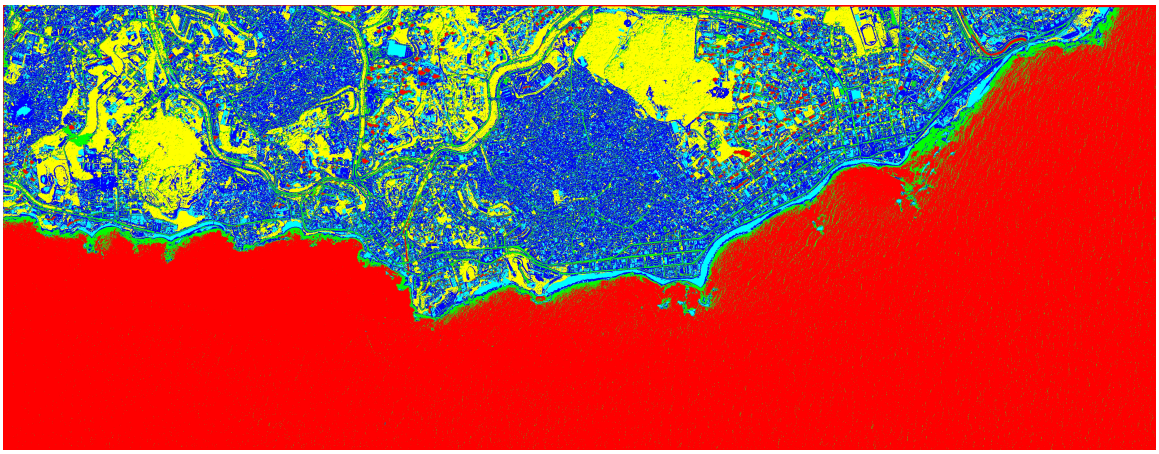
The discussions up until now have centered about unknown targets over relatively uniform terrain. There are sometimes cases that arise in which while specific targets of interest are not immediately known, the type of terrain over which they might be expected to be found is. An example of this can be illustrated with a coastal scene (Fig. 2.2b), where the surveillance and monitoring of water borne vehicles could be of interest. Processing an image that contains inland portions of developed urban regions and thresholding the results would only cue toward the those areas. From strictly tile area material complexity

perspective, this result is as accurate since the urban areas will have a greater abundance of variety of materials compared to that over the water region. This straight forward computed metric however, may not be best suited for the specified task, leading the researcher to draw influence from the body of research that viewing patterns of a scene are task influenced in what is referred to as a top down approach (Yarbus, Haigh, & Rigss, 1967).

If one considers visual saliency and areas of high contrast as natural attractors of the human visual system from a bottom up feature based model, the large area complexity metric can be thought of as the beginnings of a saliency model for multi or hyper-spectral images for use in computer vision applications. Since the value of the returned complexity metric is dependent on the individual spectrum at each pixel location, one computes the complexity map as one would normally. To investigate if the algorithm can be improved to be more task driven with some a priori processing the researcher performed an unsupervised k-means segmentation of the image for five classes (Fig 2.9a). The researcher then proceeded to filter the noise using the 3x3 median filter, before isolating the class which corresponded to the water to create a binary mask to multiply with the results of the normally computed map (Fig 2.9b).

As each tile is associated with a single metric, one could then take the mean of the pixel values of each tile to associate with the centroid coordinate and assign it as the complexity metric. Under this model, if the tile was completely classified as water, its complexity metric would remain unchanged from what would be normally computed, while those that were not would be reduced to zero. Tiles which contained a mix of water and non-water classified pixels would have their values averaged down from the straightforward computed returned result.

Comparing the top 5 percentile of thresholded metrics between the unmasked (Fig 2.10a) and masked (Fig 2.10b) results visualized by the luminance of the individual tiles demon-

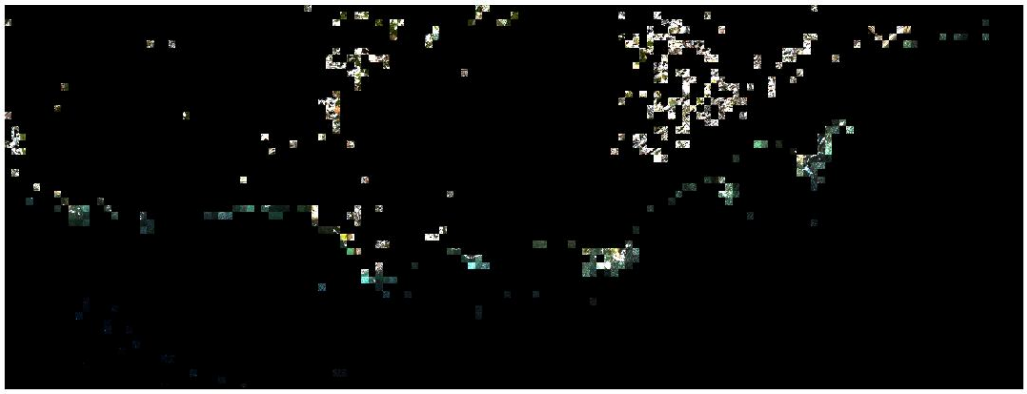


(a) K-Means Classified Image of Salvador, Brazil using 5 Classes

(b) Water Masked Complexity Metric Calculation

Figure 2.9: Two Step Processing of Complexity Map Using a Binary Class Map

(a) Standard Complexity Metric Calculation



(b) Water Masked Complexity Metric Calculation

Figure 2.10: Thresholded Complexity Results Visualization (Contrast Enhanced)

strated an improvement of identifying features of interest in the water. Due to the higher values associated with the mistakenly classified pixels of the urban areas in the top right quadrant of the image, the recomputed complexity map continued to return the peak values for the metric, despite being averaged down to a lower value. The difference was most noticeable in the edge of the coastal region as well as some tile features in the water itself as opposed to the standard method which was dominated by the urban regions. The

demonstration holds promise for the implementation of more sophisticated segmentation algorithms to yield further improvements using much more sophisticated image segmentation algorithms.

This concept can be extended further in future implementations such that the underlying expert gaze simulation not just based on a Gramian calculated volume of the convex hull defined by end-member selection but as a function of multiple measures. By incorporating techniques that mask for specific terrain coverage, and computing feature based searches such as edge detection or Haralick textural features (Haralick, Shanmugam, & Dinstein, 1973) of each subtile a new weighted metric could be developed. The researcher believes that this could technique would not only benefit improved identification of cue locations, but potentially be used to model spectral saliency for computer vision applications.

2.4 Subtle Gaze Direction

The subtle gaze direction leverages mechanisms of peripheral vision outside of the foveal view and its excellent ability to detect motion and capability to detect flicker stimuli and that a stimulus supra-threshold in color, contrast, size and temporal modulation can trigger a saccade in its direction (Lachenmayr, 2006). Using this technique of applying luminance modulations at desired cueing locations along with incorporating real time eye gaze data, the appearance of modulations can be controlled to occur only in the peripheral vision of an observer and terminated as soon as their gaze falls within the limits of the foveal view or a saccade has been made in the direction of the cue. Under ideal execution of this strategy the observer will have their gaze directed, but not experience the image modulation in their high acuity vision and minimize any perceived alteration of the stimulus (Bailey, McNamara, Sudarsanam, & Grimm, 2009). Further studies with this method of gaze manipulation were conducted to demonstrate improvements with interpretation of mammography images

(Sridharan, Bailey, McNamara, & Grimm, 2012) and embedded targets in search task performance evaluations with computer generated naturalistic scenes (McNamara, Bailey, & Grimm, 2009).

A key implementation in the mammography studies of Sridharan et al. (2012) was the ramped onset of the flicker stimulus contrast instead of a chosen fixed intensity allowing for adaptivity to individual visual sensitivity and image content differences over the defined cue duration. The implementation utilized a fixed base flicker contrast, increasing as a linearly modulated sinusoid until the cue stimulus exceeded the threshold to induce a saccade in its direction. If a fixed terminal intensity is selected, a linearly increasing modulation has the advantage of a very rapid onset but loses some fidelity (Fig. 2.11) as increasing luminance is perceived by the human visual system as an exponential function according to Weber's Law. This rapid onset was desirable given the brief amount of time over which a radiological image is scrutinized and the goal of directing an observer's gaze sequentially to mimic that of an expert of the short exposure to the image prior to assessment. The analysis of aerial imagery differs in that the images may be scrutinized over longer periods of time with intensive searches being conducted simultaneously. For this reason the framework proposed employs a longer time to maximum cue contrast of five seconds and an exponential increase in the cue intensity to prevent constant interruption to an observer's unassisted search.

The color content and dynamic range of the source image data encountered presented additional challenges. Due to the high dynamic range of image capture, select areas of the images can often exceed the image display's capabilities are rendered as saturated regions. Applying a sinusoidal modulation to a saturated or extremely dark portions of the image would result in only half the contrast differences in the midtones of the stimulus as the higher and lower intensities of the modulation reach an extreme and cannot be rendered any brighter or darker. To address these challenges, the sub-image tile to be modulated

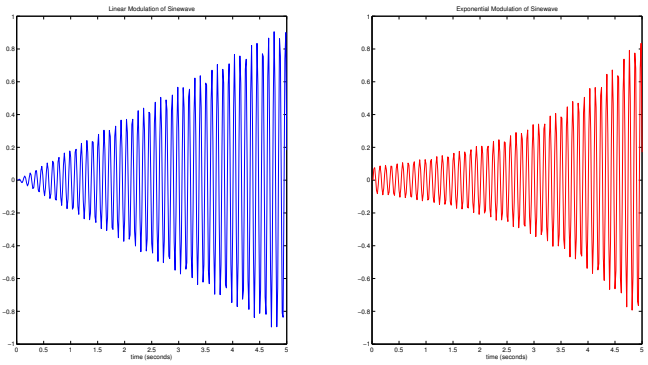


Figure 2.11: Linear vs Exponential Modulated Sinusoidal Cue

was converted to the YCbCr color space and the peak modulation intensity was set as a percentage of the mean luminance of the luma channel (Y), and this was thresholded with a minimum value for dark tiles as a first order attempt to normalize the perceived cue intensity as a function of underlying image content. This principle of image modulation formed the cueing mechanism that was employed when processing the stimulus image for display.

2.5 Stimulus Rendering and Target Cueing

Thus far the means of determining locations of potential interest and the psychophysical mechanism to direct the eye to a desired location were discussed. The implementation of the framework though, was only realized when acting on a set of logic that determined the activation, suppression and termination of a cue. This was handled by the image rendering and display component which received its inputs from an observer's real time gaze data and cue location list and acted on them based on a predetermined logic.

The first step in constructing the logic that drove our system required a basic understanding of display size and location of the observer to determine viewing geometry and

the subtended visual field, in terms of pixels per degree. This is can be computed under a geometric construct with trigonometry (Fig 2.12). For a display with horizontal and vertical viewing dimensions X and Y and an observer placed at a distance d, the subtended half angle of the field of view in the vertical (ϕ) and horizontal (θ) directions can be calculated. Doubling this and dividing it into the display resolution allows us to arrive at a mean measure of pixels per degree for image processing computation purposes.

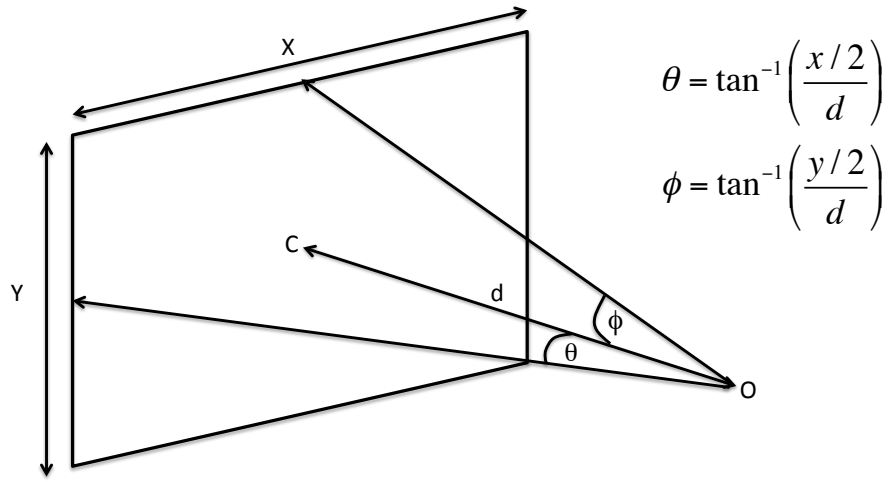


Figure 2.12: Viewer and Display Geometry

Under ideal conditions, one could select the tile sizes for the complexity metric to be calculated to correspond to that of a foveal image and the eye would be directed to the location. Accounting for reported accuracy of an eye tracker and the calibration accuracy within individual observers the researcher selected both the computed tile size and the modulated stimulus to be of 30 pixels corresponding to a 2/3 degree of visual angle. The choice of sub-foveal tile size increased the likelihood of the tile still falling within the foveal and para-foveal region of vision in the presence of calibration and accuracy uncertainties.

The logic structure used eye gaze data to determine whether it met conditions and

thresholds specified to determine how the stimulus was presented, if a cue should be triggered, and when to move to the next cue in the list of locations. This was performed by iterating through the logic shown graphically in Figure 2.13:

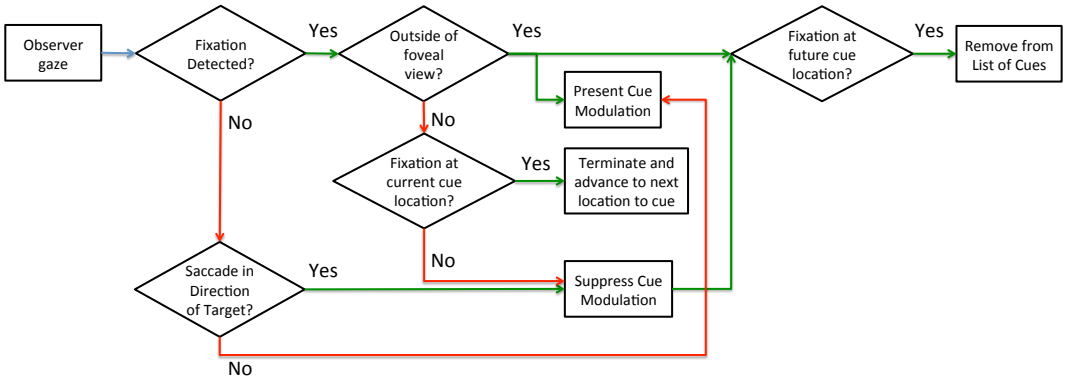


Figure 2.13: Flow Diagram of Cueing Logic

1. Determine if the observer is in the midst of a fixation
2. If deemed a fixation, determine if the point to being cued is within the foveal view location
3. If it is not a fixation, determine the direction of the saccade with respect to the cue location

The minimum distance was set such that the cue would be active while outside the foveal or para-foveal view of an observer. The determination of a fixation estimated by monitoring the range of both x and y co-ordinates not to exceed a predefined velocity in degree/second, expressed as a pixel distance over the sampling window duration. If determined to be a fixation the Euclidian distance was computed against all remaining cue locations and removed from the list if under a threshold distance and deemed to have been examined. In the cases where the samples over a given sample time exceeded the set range threshold, a

saccade was assumed and the angle from the fixation to the target location was calculated using the dot product of the vector formed by the saccade and that of the previous gaze coordinate and cue location (Fig 2.14). If the angle of the saccade was sufficiently small, the saccade is deemed to be moving toward the target and removed from the list of cueing locations.

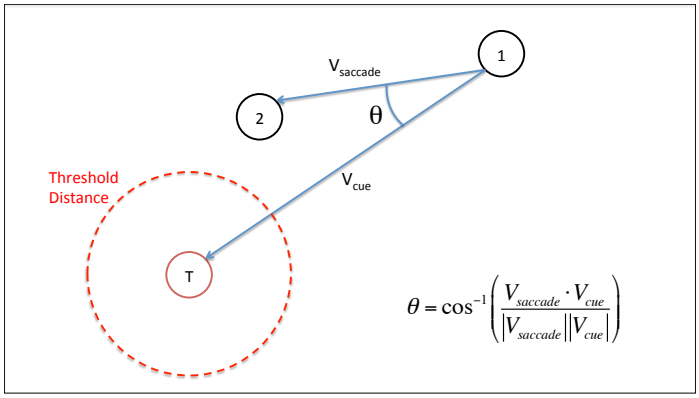


Figure 2.14: Determination of Saccadic Movement to Target

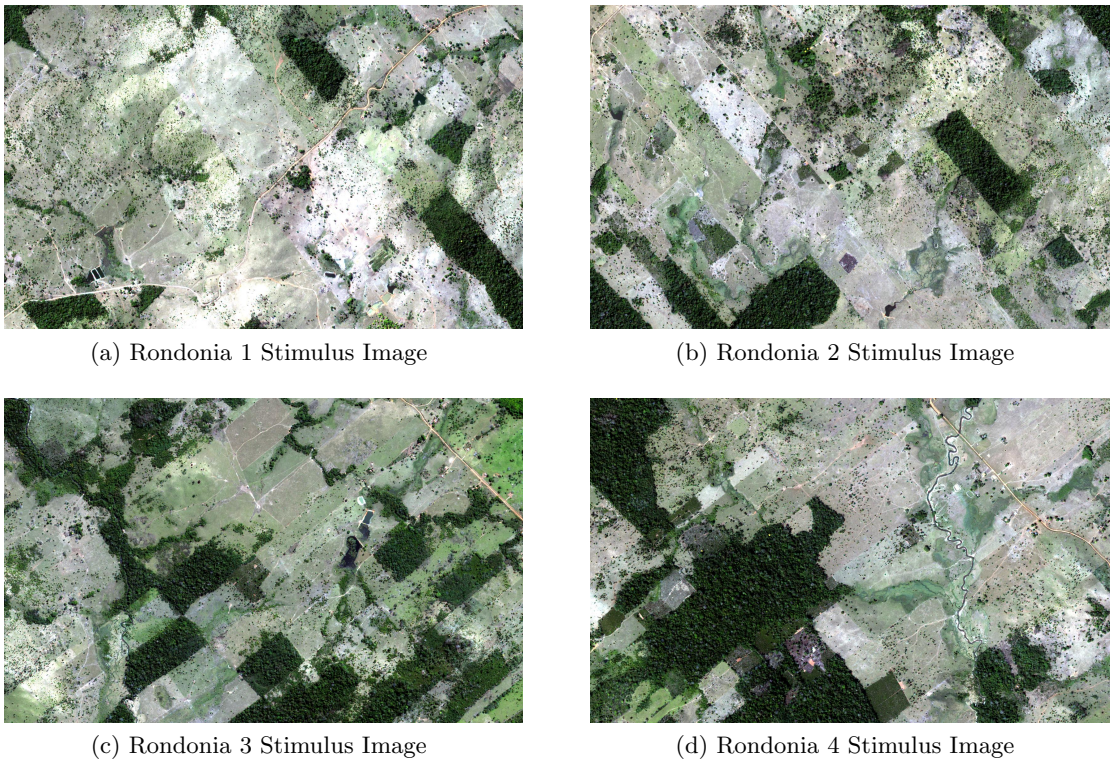
The researcher differs with this proposed implementation in that a saccade initiated in the direction of the cue was not deemed a satisfying condition to move to the next location in the list. Instead, an eye movement of an observer to attend to other areas of the image which may have been in the general direction of the active cue that may under or overshoot, but not land on the cued area was considered a distinct possibility. The researcher wanted to maximize the probability a cued area was scrutinized and terminated the cue to reinitiate it. Each cue location is subject to the logic structure for a predetermined maximal duration, or number of consecutive unsuccessful attempts before advancing to the next location so as to allow as many of the locations to be cued as possible within the allotted observer viewing time.

Chapter 3

Experimental Approach

3.1 Stimulus Images

Ten images derived from the images used to study the complexity metric in chapter 2 were used for stimulus images and presented at 1:1 scale. Each large area image mosaic was preprocessed to evaluate the large area complexity metric in the ENVI imaging software using the RIT developed spectral tools package with 30 pixel tiles and thresholded for the top 5 percentile of results to determine cueing regions across the mosaic. The stimulus images were then derived by cropping 1680x1050 pixel sub images at various orientations and rotated to disrupt spacial continuity from one image to the next where possible. Four images each were extracted from the Esperanza and Rondonia scenes, while the Salvador coastal scene was divided as two separate images totalling ten that participants were to study for a duration of 60 seconds each. Figure 3.1 shows 4 images derived from the Rondonia mosaic. All stimulus images are provided at larger resolution along with identification of targets in Appendix I.



(a) Rondonia 1 Stimulus Image

(b) Rondonia 2 Stimulus Image

(c) Rondonia 3 Stimulus Image

(d) Rondonia 4 Stimulus Image

Figure 3.1: Rondonia Image Tiles Extracted from Mosaic and used as Experiment Stimulus

3.2 Method

A sample population of 24 observers ranging from 21-32 years of age, all with normal or corrected to normal vision were selected to view and identify potential targets of interest within 10 images derived from 3 large area satellite images/mosaics. Participants were students recruited from the Rochester Institute of Technology and classified as either novice or expert using a distinguishing criterion of experts as students with study expressly focused in the area of remote sensing study or those who worked closely on projects with the remote sensing group. Half of each group were selected to be in the experimental group which were presented with cue stimulus, while the remainders were allowed to search the images freely as a control group.

The multispectral cueing framework described in Chapter 2 was implemented using a RED 250 eye tracker from SensoMotoric Instruments (SMI), sampling at a rate of 250 Hz connected to a host computer to record observer data. The display stimulus was then controlled by a second client computer running code written using Matlab and PsychToolbox (Kleiner et al., 2007), to control and render the stimulus images and cues which polls host for the SMI RED 250 data stream asymmetrically with two samples spaced at 4ms occuring between the display refresh rate of 60 Hz.

Participants were seated in a chair with fixed position to minimize body movement approximately 70cm from a display measuring 474 x 297 mm. For a display resolution of 1680 x 1050 pixels respectively, and an observer whose view is seated 70 cm away, the subtended angular field of view (FOV) is 37.41 x 23.95 degrees and corresponds to approximately 45 pixels per visual degree as calculated using the viewer geometry described. Prior to the experiment, an automated 9 point calibration using SMI's iViewX software was called remotely for each participant and validated with four locations independent of the calibration sequence locations. Participants were instructed to reject any calibrations that did not fall within 1.5 degrees of error in either the horizontal or vertical direction and repeat the process before commencing the study. The presentation of each image set was preceded by text informing the participant of the terrain cover, ground resolution of the image and specific potential targets of interest. Participants were instructed to identify potential targets of interest by hitting the space bar and that the location of their gaze would be recorded with no alteration to the image being displayed. The task given to participants was to identify building structures in the Esperanza and Rondonia image sets, and watercraft in the Salvador coastal scene. Each stimulus image was separated by a center fixation cross to ensure observers started viewing the image in the same location and also served as a periodic verification that the calibration results had not drifted too far. In the event that

a center fixation was not detected within 1.5 seconds, the system was deemed to require re-calibration and called automatically. Stimulus presentation was fixed at 60 seconds per image for all viewers with all images presented in sequential order for all subjects.

Each image was presented using the visible spectral bands and a 2% linear stretch. In the case where stimulus cues were presented a 30 pixel modulated tile was blended with the underlying stimulus using a gaussian mask with a modulation rate of 10Hz as with the original subtle gaze direction technique of Bailey et al. (2009). A minimum distance of 200 pixels was required for a cue modulation to be presented and an angular threshold of 10 degrees from the stimulus was considered a saccade in its direction. Cues were set to increase in intensity exponentially from $e^{-2.5}$ up to a maximum of the greater of 0.13 or 0.20 of the mean luminance of the region in the luma channel over a 5 second duration. A secondary condition of a maximum of three attempts to cue the observer was placed in the event saccades were made in the direction of the target but not fixated on. The zone of fixation was calculated as the region being bounded by the midpoint of the range of the current and previous 3 gaze samples and all points inclusive within a 80 pixel radius corresponding to a 1.78 degree boundary.

A brief post-study interview was conducted with each participant to specifically ask if they noticed any image display anomalies, if so whether they felt these anomalies were distracting in any way or felt that it impaired them from performing the assigned visual search task to the best of their ability. Finally there was an open invitation to provide any feedback they felt would be helpful in improving their search task.

Chapter 4

Results and Discussion

4.1 Analytical Approach

The development of a dynamic cueing system that removes cues that have been deemed fixated on that may arise later in the sequence, and suppress or advance to the next cue based on individual observer behavior can lead to efficiencies and optimize the user experience, but proves to be very challenging in its analysis as the gaze path for each participant in the experiment group may be different. The approach taken in this analysis was to look at the statistics across each individual image as well as participant, to gain some insight to the effectiveness of the cue stimulus for specific images and if there was variation between cued observer and differences in individual sensitivity. Since the specific instances in time during which a cue is active and its directed location are known, we could analyze cues of a specified duration and the distance of the series of fixations that occurred from the onset of the cue to its termination. In the analysis, the researcher also included one fixation post cue termination since the model ideally terminates the cue prior to allowing an observer to arrive at its location with their gaze. The raw data was event processed using SMI's software to distinguish between fixations, saccades and blink events. A threshold of a 50 ms

minimum fixation duration was specified for event detection with the results being plotted and verified against some raw data prior to proceeding with its use.

4.2 Analysis of Cue Efficacy Within Images

The hypothesis was that if the framework was effective at guiding observer gaze within an image, an observer's gaze would make a series of saccades and fixations toward the cue modulation. It would therefore be expected that a histogram of the distance of fixations from any given point by free observers would follow a normal distribution (given it is a random event) while those of directed observers would show a shifted normal distribution toward smaller distances. The challenge arose in how to compare the unique sequence of cues among the experiment group in time with those of the control group to determine whether cues were effective at guiding observer gaze.

It was expected that the free and undirected gaze of an undirected observer was random in nature and the distance from any given cue should also be random and uniformly distributed from any given cue location. The directed sequence may be different for each observer in the experiment group due to the dynamic nature of cue selection and removal of cues deemed to have been fixated on by a participant by the logic and cueing module. This meant it was not a straightforward task of comparing the specific instances in time of an undirected observer with that of the experiment group. Recall also, that the cues were designed to increase until reaching above threshold values for a just noticeable difference to induce a saccade in its direction. Based on this behavior if one were to analyze fixations across the time window over which the cue was present, one would expect cues of longer duration to have increasing fixations away from the target until it reached its intended effect and for a saccade and fixation toward the cued location. In the analysis, the researcher sought to neutralize this effect, by analyzing one fixation after the cue terminated and 600

ms prior for a fixed window. The window size was selected to capture 2-3 fixations prior the last fixation and with the estimate that a typical fixation lasts 250 ms.

For the comparison to the control group the researcher performed random permutations and paired the control group and compared that the instances in time of specific cue durations in a trial of cued observers as shown graphically in Figure 4.1. Using this Monte-Carlo type simulation, the histograms of fixation distances expressed as a percentage for visual comparison and the cumulative number of fixations within 5 degrees for short, medium and long cue durations were reported. Short, medium and long cue durations were defined as cues lasting between 500-2000, 2001-3500, and 3501-5001 miliseconds respectively. Cues lasting less than 500 ms were excluded as the stimulus is negligible at initial onset and required time to increase to supra threshold intensity and saccades themselves take 150-200 ms to plan and execute.

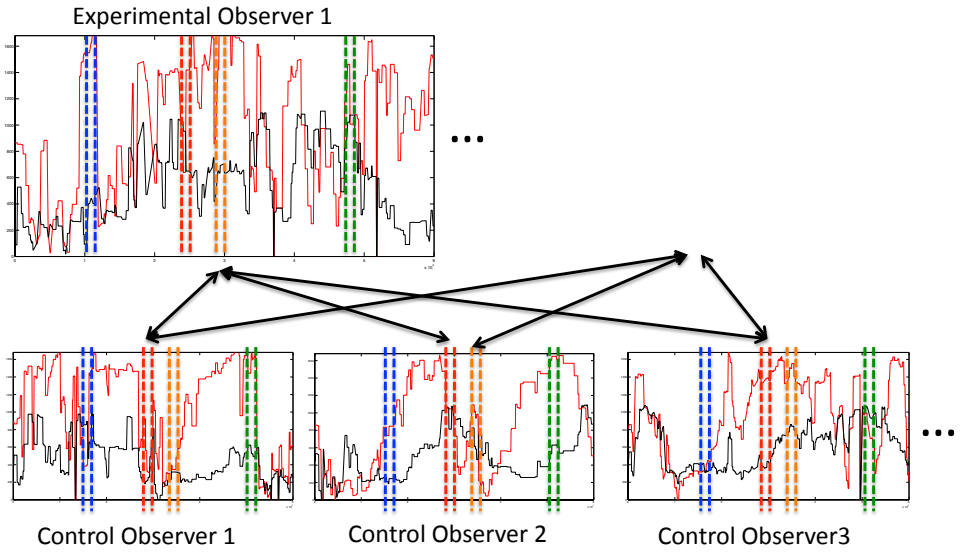


Figure 4.1: Graphical Depiction of Comparisons Between Control Group to Random Permutations of Experiment Group

It was felt that any fixations that happen to land in the fixation zone in less than this amount of time was considered a random event. Figure 4.2 presents a visualization of histogram of fixations at a given distance from a cue expressed as a visual angle for cued observers and a Monte-Carlo simulation as previously described with 20 iterations using the data from non-cued observers.

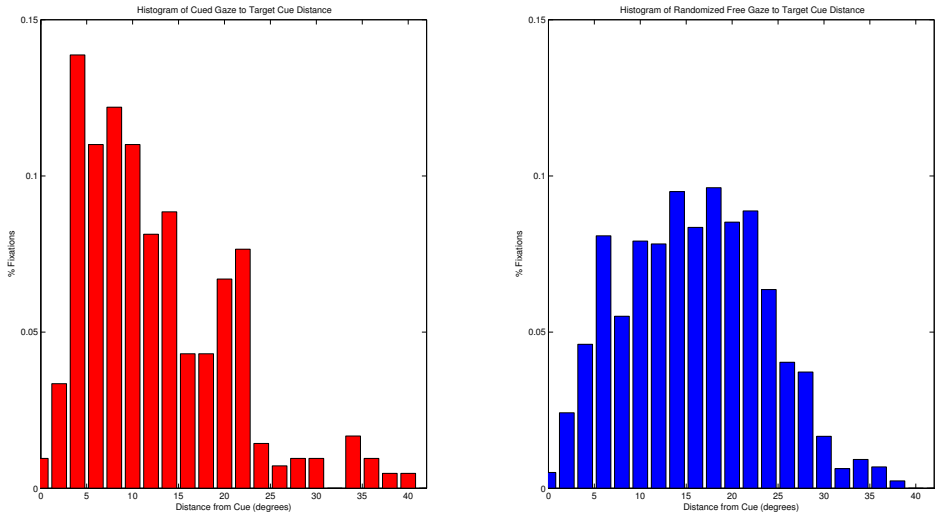


Figure 4.2: Histogram distribution of fixation distances between cued and free gaze observers for cues of 500-2000 ms durations for the Rondonia 4 Image

The histogram shown demonstrated a strong indication that cue stimuli were effective for the given image. Using the measure of cumulate fixations within 5 degrees, a threshold chosen to account for the combined calibration and tracking uncertainty estimated at 1.5 degrees, and the angular foveal view of 2 degrees, the researcher felt confident that observers within this set threshold had the intended location of gaze within their foveal or para-foveal view. For each image, the metrics of cumulative fixations within 5 degrees and skewness as defined by equation 2.4 were calculated. For the histogram of the image chosen (Rondonia 4), cumulative fixations of all cues falling within 5 degrees of the target of cued observers

was 20.33% versus 9.13% of undirected observers. Skewness of the distribution was also calculated at 1.0284 and 0.1865 for cued and non-cued observers respectively. The metrics for each image were calculated and tabulated in Table 4.1-4.3 for short, medium and long cue durations along with the all the plotted histograms included in Appendix II .

Table 4.1: Fixation Histogram Metrics for Cue Stimuli of 500-2000 ms

Image	% Cumulative Fixations $< 5°$		Skewness	
	Cued	Non-Cued	Cued	Non-Cued
Esperanza 1	11.65	7.45	0.1950	0.2820
Esperanza 2	11.01	5.71	0.2468	0.2790
Esperanza 3	13.27	11.90	0.3875	0.0527
Esperanza 4	28.75	26.33	0.9956	0.6729
Rondonia 1	21.88	10.52	0.5972	0.0249
Rondonia 2	17.99	10.65	0.8488	0.6423
Rondonia 3	9.50	7.52	0.2592	0.1256
Rondonia 4	20.33	9.13	1.0284	0.1865
Salvador 1	18.04	13.64	0.7097	0.3122
Salvador 2	19.33	15.74	0.8707	0.6347

No specific trend between the percentage of cumulative fixations that fall within 5 degrees of the target and skewness that is calculated was observed. In general, the observations showed that cued participants tended to have a higher percentage of fixations within the threshold and a higher skew in the histogram as well, in all except two cases (Esperanza 1 and Esperanza 2) when compared the non-cued observers. This suggests there was efficacy to the cueing stimulus. As the cues of greater duration were examined, there was less marked difference in the number of fixations between cued and non-cued observers as summarized in the Tables 4.2 and 4.3.

It was noted that the Salvador images in general did not appear to exhibit any difference in the fixations within threshold. This was somewhat expected as even with the water class masking, the improved results for the coast and water still fell within the low end of the

Table 4.2: Fixation Histogram Metrics for Cue Stimuli of 2001-3500 ms

Image	% Cumulative Fixations < 5°		Skewness	
	Cued	Non-Cued	Cued	Non-Cued
Esperanza 1	12.36	11.04	0.2566	-0.0973
Esperanza 2	4.61	4.65	0.6212	0.4327
Esperanza 3	17.37	11.18	0.4424	0.2122
Esperanza 4	19.23	21.12	0.8290	0.5007
Rondonia 1	9.89	8.93	0.2996	0.0309
Rondonia 2	10.00	9.31	0.3544	0.6622
Rondonia 3	6.96	7.49	0.0151	0.0024
Rondonia 4	17.26	8.97	0.4222	0.1529
Salvador 1	13.82	10.34	0.5052	0.2409
Salvador 2	8.70	14.59	0.1504	0.5198

Table 4.3: Fixation Histogram Metrics for Cue Stimuli of 3501-5000 ms

Image	% Cumulative Fixations < 5°		Skewness	
	Cued	Non-Cued	Cued	Non-Cued
Esperanza 1	8.33	9.57	-0.0565	0.2466
Esperanza 2	7.86	6.04	0.3189	0.4407
Esperanza 3	13.74	8.23	0.4357	0.0607
Esperanza 4	23.27	15.85	1.0241	0.5883
Rondonia 1	10.46	8.85	-0.1250	0.0088
Rondonia 2	18.26	8.97	0.3727	0.5404
Rondonia 3	17.33	9.21	0.5908	0.2808
Rondonia 4	21.43	4.90	0.3608	0.2087
Salvador 1	12.32	15.55	0.6154	0.2378
Salvador 2	14.61	16.07	0.5336	0.4185

threshold with the urban areas dominating and being at the top of the list of areas to be cued first. Since the task inherently directed a viewer to instinctively focus their attention to the coast and water regions of the image, the images were included in the study for thoroughness to see if irrelevant cues would present a distractor from the assigned task. The tabulated measure for the Salvador images supported that the presence of the cue did not detract from the assigned task and only one participant reported noticing any flickering in the images but did not feel it impaired them from performing the task.

4.3 Visual Search Performance

Demonstration of the ability to guide the eye to areas of interest within an image was an important first step towards developing a useful framework for visual search. However the researchers were also interested in knowing if this guidance was able to provide any sort of performance benefit. One way to do this is through traditional signal detection models and assume the signal and noise are normally distributed and model the distance between distributions based on the hit and false alarm rates of targets by observers. The researchers efforts to measure this presented a number of challenges. As the source images exceeded thousands of square kilometers, establishing ground truth in the images was not a trivial task. To this end, we acknowledged that the researcher acted as an omni-sentient observer prior to the study and performed an exhaustive visual search using higher resolution panchromatic images to classify what were to be deemed as targets within the stimulus images which introduced an element of uncertainty in and of itself.

Another challenge arose from the fact that observers were asked to identify targets with a key press and the location was determined through sifting through and analysis of the visual tracking which is known to have uncertainties introduced by both the calibration and accuracy of the tracker itself. Therefore, one is not able to simply take the raw gaze

co-ordinate at the point of the key press at the time of target identification. Instead, the researcher used the time stamp from the key presses and performed a search against the event detected fixation data to find the fixation during, or just prior to the key press in the case when it was during a saccade, and applied a 3 degree radius around the co-ordinate. If a designated target(s) fell within this area, it was deemed a 'hit' and in cases where no targets existed within the bounds considered a 'false alarm'. In the event the key press occurred during a blink or track loss, the data point was not used in our analysis as there is no way of knowing where the observer was looking at time of identification and would bias the result. Using the mean hit and false alarm rates as indicated, if a normally distribution of signal (target present) and noise (background) are assumed, then the hit and false alarm rates can only be achieved by a fixed distance between the Gaussian functions. The distance between the peaks of the normally distributed signal and noise found by taking the difference of the inverse cumulative Gaussian distributions results in the metric d-prime, which can be interpreted as a measure of ease in distinguishing between a target present vs background only condition. A first order measure of d-prime was assessed to look for any observable differences in performance between cued and non-cued observers for each of the images. The Esperanza 2 Image was excluded from this analysis as it did not have any targets in the image and would only have resulted in false alarms.

Table 4.4 summarizes the mean hit and false alarm rates of the cued experiment group and non-cued control groups for each image. The hit rate was calculated to be number of targets correctly identified over all deemed targets in an image, while the false alarm rate was the number of incorrectly identified targets over all instances of target identifications by an observer.

While the analysis of cue efficacy in the previous section appeared to demonstrate the greatest success at drawing the eye to the cues in the Rondonia Images, the performance

Table 4.4: Mean Hit and False Alarm Rates Between Cued and Non-Cued Observers

Image	Hits (%)		False Alarms (%)		D-prime	
	Cued	Non-Cued	Cued	Non-Cued	Cued	Non-Cued
Esperanza 1	67.34	65.66	9.42	19.80	1.7647	1.2521
Esperanza 3	13.33	13.89	41.67	73.02	-0.9003	-1.6988
Esperanza 4	63.23	51.11	11.69	26.37	1.5287	0.6598
Rondonia 1	32.46	45.45	28.99	27.89	0.9987	0.4718
Rondonia 2	26.26	57.58	34.56	26.34	-0.2380	0.8240
Rondonia 3	50.30	55.15	26.91	33.35	0.6233	0.5596
Rondonia 4	57.44	61.98	32.25	30.73	0.6484	0.8086
Salvador 1	48.48	66.67	26.74	25.56	0.5827	1.0876
Salvador 2	43.18	48.66	68.18	42.57	-0,1382	0.6600

of target identification performance did not coincide to cueing efficacy. The most noted improvement in performance noted is in the reduced false alarms between 75% or greater compared to the non-cued case. While one would expect effective cueing would increase performance through aiding in the detection of targets, we did not see a marked improvement in positive identification of targets or 'hits'. One hypothesis to explain this result is that drawing the eye away from other areas that might trigger false alarms reduces their occurrence. The Esperanza data set was also arguably the most foreign in appearance to observers and it could have been that despite the lower success rates of the cue stimulus that its cues had a greater suggestive effect, but the very modest improvements does not seem to support this. Based on these results alone and the small sample size, it was not possible to draw any conclusions on whether the cues directly impacted target search performance.

4.4 Qualitative Observations

After completing the study, all participants were asked if at any point they noticed any screen refresh issues or any kind of display anomalies. No participants in the control group

reported noticing any issues while 8 of 12 in the experiment group reported noticing intermittent screen anomalies during several images. Of the 8 participants that reported noticing signs of anomalies, one participant reported it as a minor distraction to their assigned search task while another noted that the draw of the eye away from their search task used up some of the allotted time. The remainder reported recalling a flicker as being visible no more than on 6 occasions, with one cued observer citing that they found it helpful when noticed.

Chapter 5

Conclusions and Recommendations

5.1 Conclusion

The primary objective of developing an eye movement based display system that integrates spectral algorithm processing into a dynamic cueing framework was implemented and realized in the course of conducting the study summarized in this thesis. The augmentative framework demonstrated efficacy in drawing a viewer's gaze to regions of an image using the underlying spectral processing algorithm. The list of cues presented to each participant in the experiment group also dynamically altered based on individual viewing patterns and was not subject to a fixed sequence in contrast to prior implementations of gaze manipulation.

The analysis of fixations just prior to the cessation of cues indicate the parameters selected for the maximum modulation intensity and its progressive onset, while the cue was active, intended to just exceed thresholds to induce a saccade were not always enough to solicit observer gaze. It is hypothesized that this may be related to the nature of the attentional load of the type of search task being conducted and that the stimulus required to induce an eye movement away from an active search needs specific psychophysical experiments to help better set their intensity in future implementations.

The analysis of fixation distances of a fixed temporal range of 600 ms extending backward from cue termination suggest that the most effective cues caught participants attention shortly after onset. This is indicated by the largest differences in percentage of fixations falling within the 5 degree threshold and more dramatic skew in the histogram shapes for the cues of shorter duration.

Improvements in search task performance observed between the experiment and control group were not conclusive given the methodology used to assess d-prime and the necessary assumptions that were made in the calculation of its measure. Furthermore, the sample size was not adequate for results to be generalized, and studies with a much larger sample size need to be conducted before any generalizations can be made with respect to visual search task performance.

5.2 Recommendations for Future Work

5.2.1 Stimulus Strength Normalization

While the framework was able to demonstrate that one can direct viewer gaze using a luminance modulation in the region of interest, it was only able to bring fixation termination of the cue landing within 5 degrees of visual arc approximate one third of the time. It is hypothesized that the cue efficacy is in fact a combination of parameters and can be expressed as a function of distance from present viewer gaze, contrast, spatial size and temporal frequency of the modulation, and potentially nature of the underlying and surrounding image content. Cue efficacy as it relates to these parameters is not well understood and we submit that a series of psychophysical experiments to study the sensitivity of each of these parameters to model a normalized cue stimulus and have a predictable time from onset to induced eye movement to be valuable for future implementations of subtle gaze direction for systems of any nature. We believe that it would also be of value to analyze the content

of the underlying images in the frequency domain as a means of studying how underlying and surrounding content may impact cue efficacy.

5.2.2 Search Task Performance Evaluation

The study conducted utilized images of very large geographic coverage and search tasks involving multiple targets. An experiment design involving single target images with a forced choice response to measure correct target identification and false alarms would be a more suitable method of accurately measuring d-prime between control and experimental groups. In addition to this, response time between cued and non-cued observers can also be taken as a measure of whether cueing improves search task efficiency.

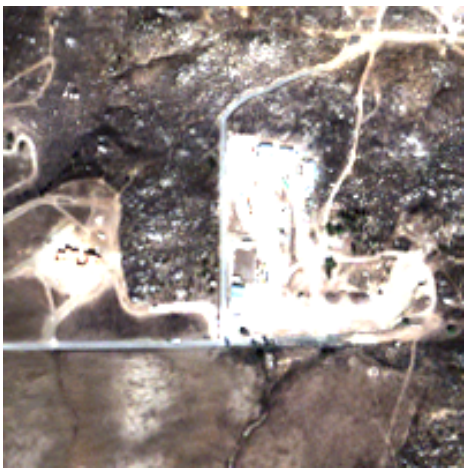
5.2.3 Adaptive Angular Thresholding

For the realization of the framework, a fixed visual angle threshold for various parameters within the system was specified such as determination of when an area had been deemed fixated on and examined, or what was considered an acceptable radial boundary before the modulation cue might become apparent. To further optimize the system for individual differences, these parameters should be adaptive based on individual calibration results with the eye tracker as well as using results from the psychophysical experiments mentioned above so that the system can impose tighter tolerances for those with a steady eye track and conversely looser criteria in the presence of high amount of noise and aid in statistical evaluation of system performance and efficacy.

5.2.4 Local Foveal Image Enhancement

The framework and display system differs from traditional gaze contingent models in that it seeks to alter and modify regions of the image away from where the observer's gaze is,

hence the description of it being an eye movement based display. The image rendering and processing engine is however, scalable and it is believed that the option of implementing local image enhancement based on the location of an observer gaze and adding this to framework can also lead to search task performance. Consider that all stimulus images presented were scaled to a 2% linear stretch based on the contents of the entire image. Given the discussions that it is only within the 1-2 degree angular field of view that the human visual system experiences high acuity in vision, it stands to reason that the foveal image should be optimized for information driven tasks and not be compromised for rendering a larger image.



(a) 2% Linear Contrast Enhacement Using Entire Image

(b) 2% Linear Contrast Enhacement Using 50x50 pixel region

Figure 5.1: Comparison of Image Enhancement Based on Global and Localized Regions

To demonstrate the potential benefits of localized rendering, we extracted a 200x200 pixel saturated region, a result from the compromise of preserving the details in the larger image to that of one that has been contrast adjusted based on a 50x50 pixel sampling region (Fig 5.1), corresponding to just over 1 degree of visual arc for the demonstrated system. This type of image processing over much smaller regions is more computationally efficient

and can yield improvements in information driven tasks. We expect that this would not create a load that would preclude the display to continue refreshing at a rate of 60 Hz and could be implemented in a separate study or in conjunction with the gaze directing framework in future work.

5.2.5 Investigation of Bias Effects and Comparison Against Overt Cueing

This exploratory research was primarily focused on the ability to direct observer gaze through the use of an underlying automated image analysis and whether this yielded any improvements in visual search tasks. Although the cueing was direct, in that it attempted to draw an observer's gaze to a specific location within the image, the motivation for using non-overt cueing would be to minimize bias. Future research to compare the effect of cueing overtly or through subtle means and how this affects operator behavior may bear evidence to which conditions it may be most beneficial to employ this type of framework.

5.2.6 Interactive Image Exploration

For our exploratory work of cue efficacy and search guidance being driven by an underlying algorithm, navigating the images within the display was not a primary concern and beyond the scope of this study. Instead of a static 1:1 pixel resolution display, future iterations of the system may be developed to allow scrolling and zooming of the window and a dynamic adjustment of precomputed complexity metrics of different tile sizes appropriate for each scale of view, and initiating cues only in the observable region of the image at any given time.

5.2.7 Comparison of Spectral Complexity to Saliency Maps as a Predictor of Gaze in Aerial Imagery

As an alternate consideration, it is proposed that one could see how well the large area complexity metric correlates with an undirected search of an aerial image. The approach in this new domain was to seek ways of augmenting observer search capability. An interesting area for exploration is to compare the large area complexity metric with that of the undirected gaze of expert image analysts. Most research in the area of visual saliency has been performed through naturalistic images and it is not known if this generalizes to aerial imagery or visual representations of scenes under geometric transformations and dimensional reduction. It is noted that it would be interesting to perform comparisons of a viewer's gaze with that of a heat map derived from the spectral complexity and current models of visual saliency maps to see if the large area complexity metric may be extended as a model for expert image analyst gaze and be implemented as part of a multi-step image processing chain as a validation model as a human observer would.

References

Andriole, K. P., Wolfe, J. M., Khorasani, R., Treves, S. T., Getty, D. J., Jacobson, F. L., ... Seltzer, S. E. (2011, MAY). Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section CT Scan Can Ruin Your Whole Day. *RADIOLOGY*, *259*(2), 346-362. doi: 10.1148/radiol.11091276

Bailey, R., McNamara, A., Sudarsanam, N., & Grimm, C. (2009, September). Subtle gaze direction. *ACM Trans. Graph.*, *28*(4), 100:1–100:14. Retrieved from `http://doi.acm.org/10.1145/1559755.1559757` doi: 10.1145/1559755.1559757

Estes, J. E., & Thorley, G. A. (1983). Manual of remote sensing, volume 2, interpretation and applications. *Falls Church, VA, American Society of Photogrammetry, 1983, 1240 p.*

Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized "satisfaction of search": Adverse influences on dual-target search accuracy. *Journal of Experimental Psychology: Applied*, *16*(1), 60-71. (http://www.ncbi.nlm.nih.gov/pubmed/20350044)

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*(6), 610–621.

Jean, G. V. (2011). Broadcast television tools to help intelligence analysts wade through data. *National Defense*, *95*(688), 32-33.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). Whats

new in psychtoolbox-3. *Perception*, *36*(14), 1–1.

Lachenmayr, B. (2006). Visual field and road traffic. how does peripheral vision function?]. *Der Ophthalmologe: Zeitschrift der Deutschen Ophthalmologischen Gesellschaft*, *103*(5), 373.

L.J, S., K.L, M., & M, B. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991-991.

McNamara, A., Bailey, R., & Grimm, C. (2009). Search task performance using subtle gaze direction with the presence of distractions. *ACM Transactions on Applied Perception (TAP)*, *6*(3), 17.

Messinger, D. W., Ziemann, A., Basener, B., & Schlamm, A. (2012). Metrics of spectral image complexity with application to large area search. *Optical Engineering*, *51*(3), 036201–1.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human factors*, *46*(2), 196-204.

Parasuraman, R., & Riley, V. (1997, JUN). Humans and Automation: Use, misuse, disuse, abuse [Review]. *HUMAN FACTORS*, *39*(2), 230-253.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, *41*(25), 3587–3596.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (1996). Modeling saccadic targeting in visual search. *Advances in neural information processing systems*, 830–836.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements in visual cognition: a computational study. *Urbana*, *51*, 61801.

Rice, S., & McCarley, J. S. (2011). Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of experimental*

psychology. Applied, *17*(4), 320-331.

Rogers, L. F. (2000). Keep looking: satisfaction of search. *AJR. American journal of roentgenology*, *175*(2), 287-287. (http://www.ncbi.nlm.nih.gov/pubmed/10915658)

Schiller, P. H., Kendall, G. L., Slocum, W. M., & Tehovnik, E. J. (2008). Conditions that alter saccadic eye movement latencies and affect target choice to visual stimuli and to electrical stimulation of area v1 in the monkey. *Vis Neurosci*, *25*, 661–673.

Schott, J. R. (2007). *Remote sensing: the image chain approach* (2nd ed.). Oxford University Press.

Sridharan, S., Bailey, R., McNamara, A., & Grimm, C. (2011). Subtle gaze manipulation for improved mammography training. In *Proceedings of the acm siggraph symposium on applied perception in graphics and visualization* (pp. 112–112). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2077451.2077475` doi: 10.1145/2077451.2077475

Sridharan, S., Bailey, R., McNamara, A., & Grimm, C. (2012). Subtle gaze manipulation for improved mammography training. In *Proceedings of the symposium on eye tracking research and applications* (pp. 75–82).

Vine, S. J., Masters, R. S., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery*, *152*(1), 32 - 40. Retrieved from `http://www.sciencedirect.com/science/article/pii/S003960601200044X` doi: http://dx.doi.org/10.1016/j.surg.2012.02.002

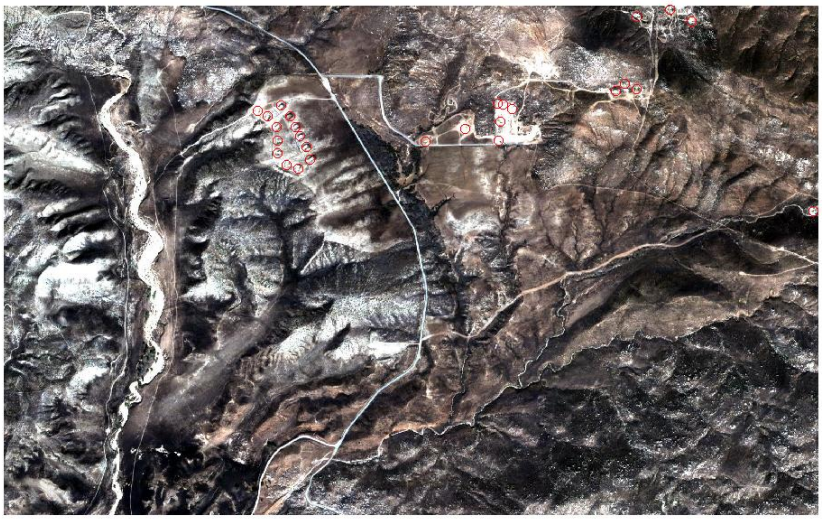
Wiczorek, R., Meyer, J., & Torsten, G. (2012). *On the relation between reliance and compliance in an aided visual scanning task* (Vol. 56) (No. 1).

Yarbus, A. L., Haigh, B., & Rigss, L. A. (1967). *Eye movements and vision* (Vol. 2) (No. 5.10). Plenum press New York.

Yeh, M., Wickens, C. D., & F, J. S. (1999). Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human factors*, *41*(4), 524-524.

Appendix I: Stimulus Image Tiles

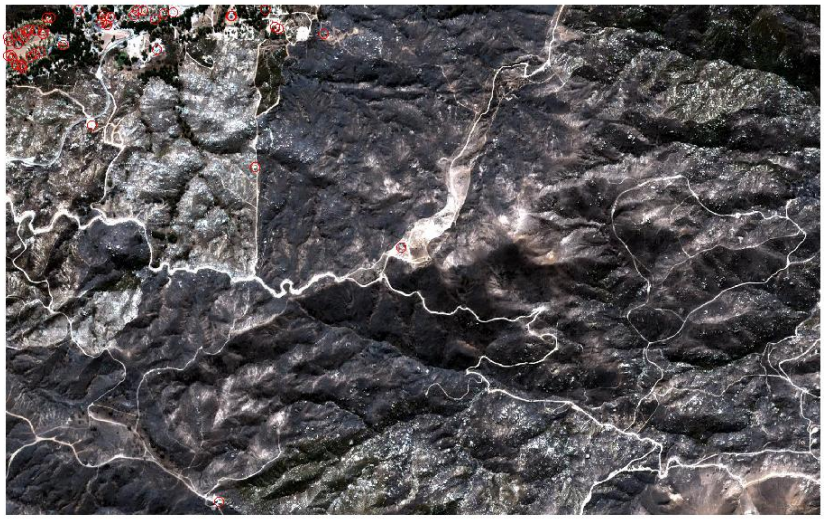
Stimulus Image Sets with Designated Targets Circled in Red



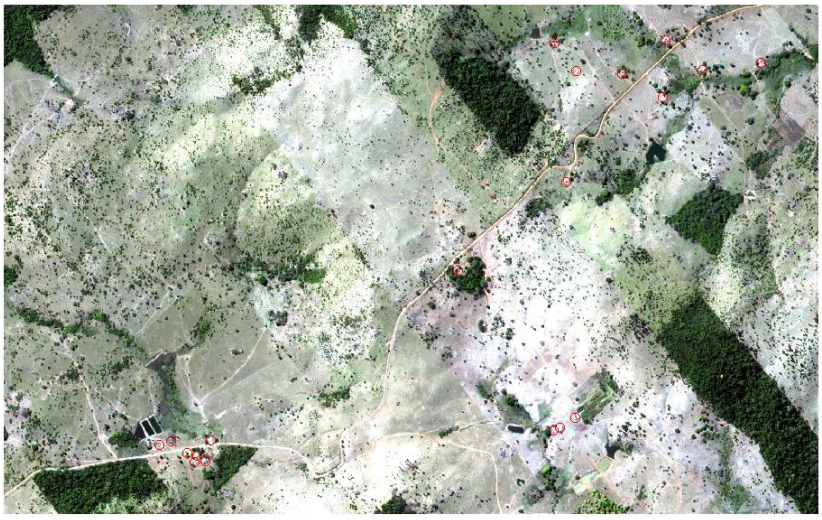
Esperanza 1 Stimulus Image

Esperanza 2 Stimulus Image

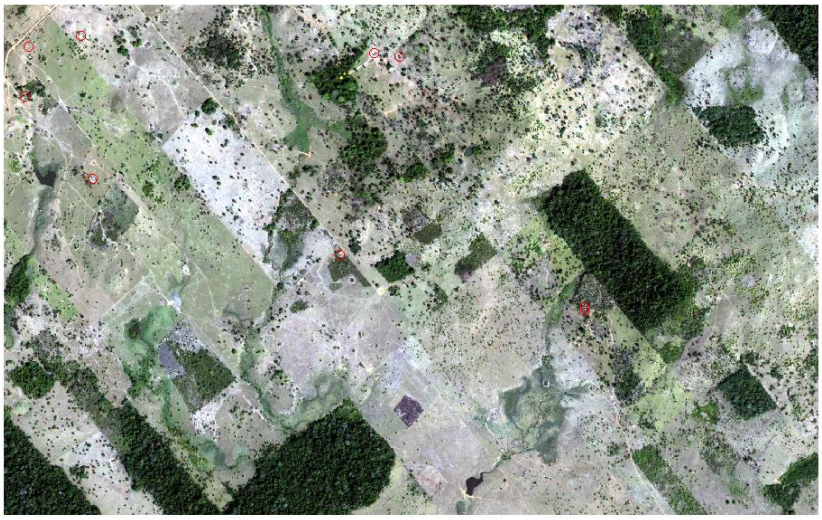
Esperanza 3 Stimulus Image



Esperanza 4 Stimulus Image



Rondonia 1 Stimulus Image



Rondonia 2 Stimulus Image

Rondonia 3 Stimulus Image

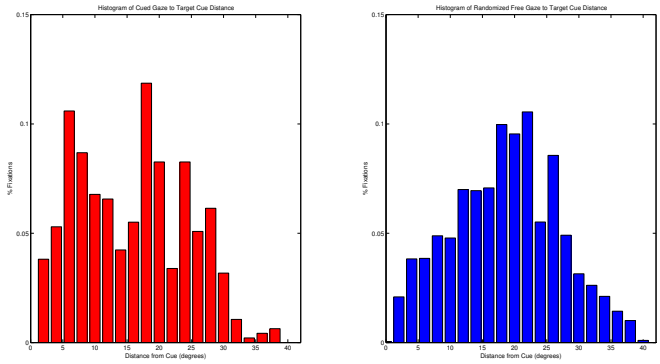
Rondonia 4 Stimulus Image

Salvador 1 Stimulus Image

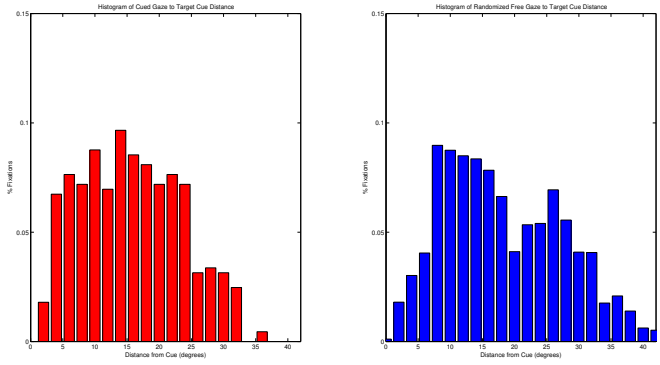
Salvador 2 Stimulus Image

Appendix II: Fixation Histograms

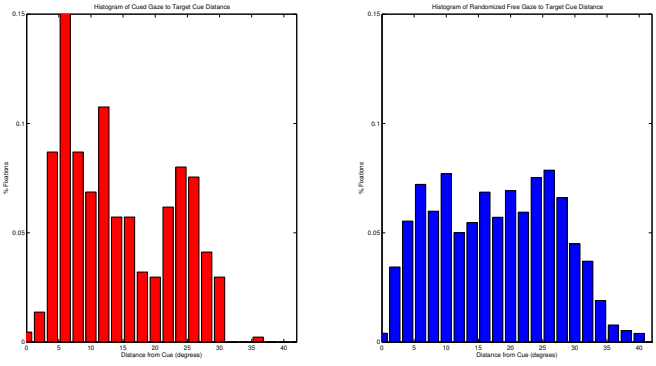
Histograms for Cue Durations of 500-2000ms



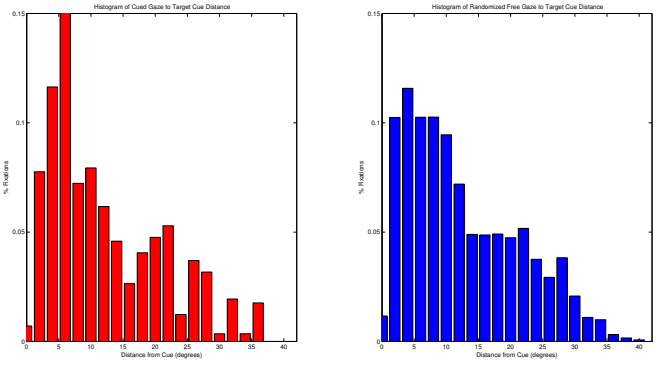
Histogram distribution of Esperanza 1 Image



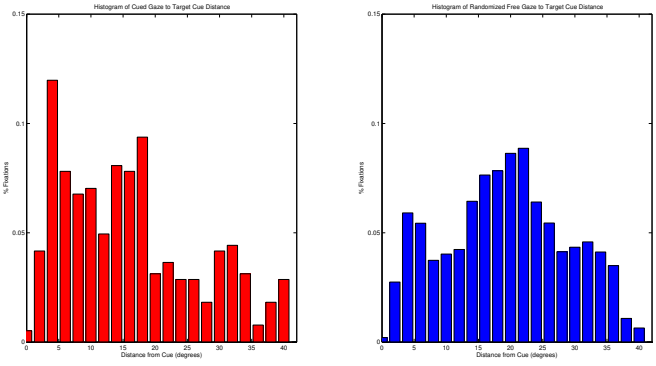
Histogram distribution of Esperanza 2 Image



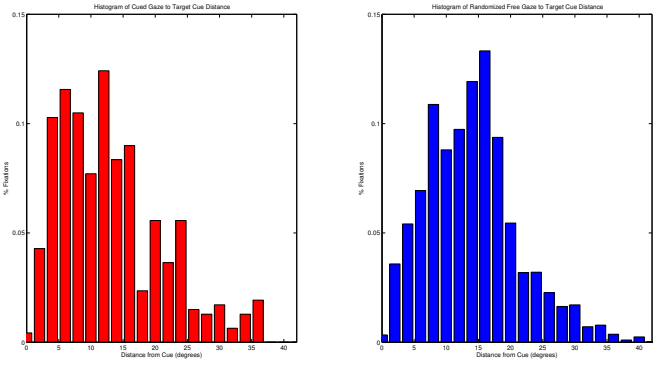
Histogram distribution of Esperanza 3 Image



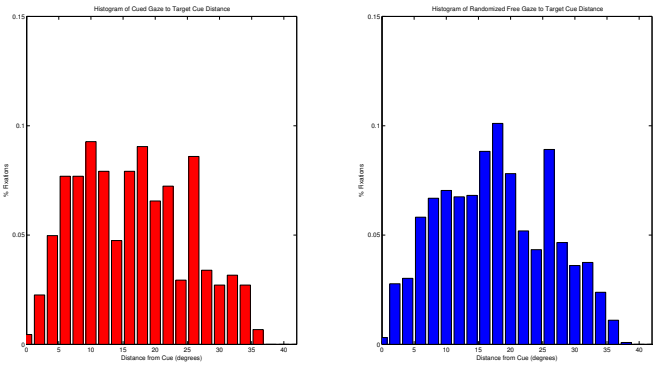
Histogram distribution of Esperanza 4 Image



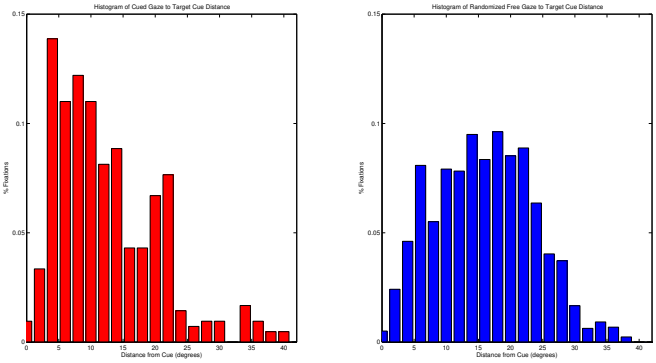
Histogram distribution of Rondonia 1 Image



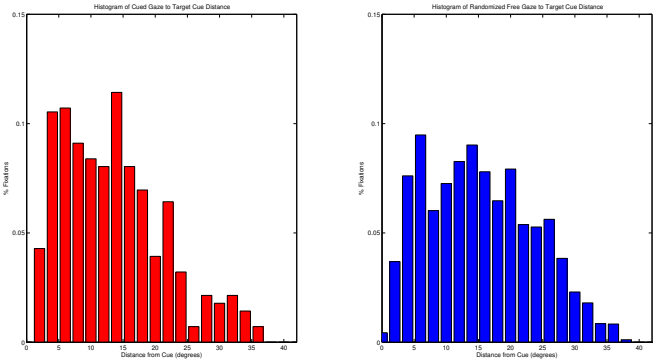
Histogram distribution of Rondonia 2 Image



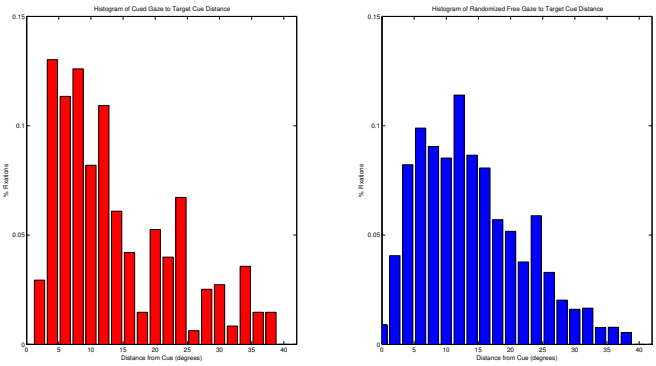
Histogram distribution of Rondonia 3 Image



Histogram distribution of Rondonia 4 Image

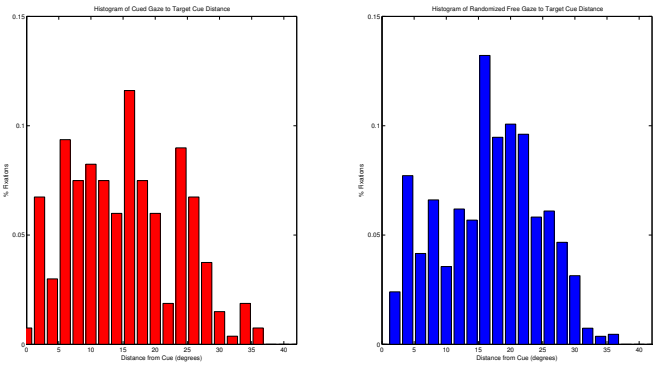


Histogram distribution of Salvador 1 Image

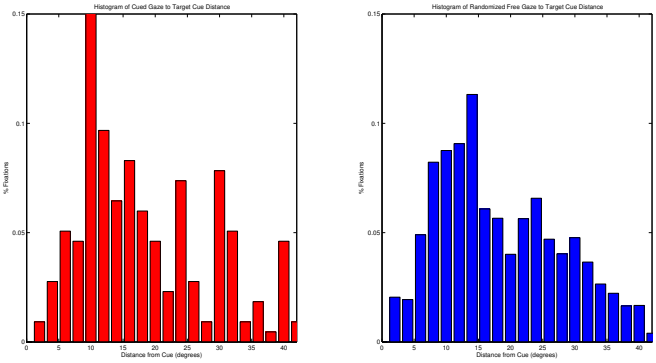


Histogram distribution of Salvador 2 Image

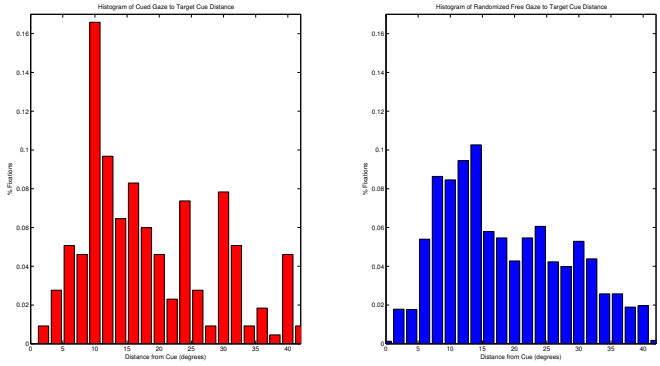
Histograms for Cue Durations of 2001-3500ms



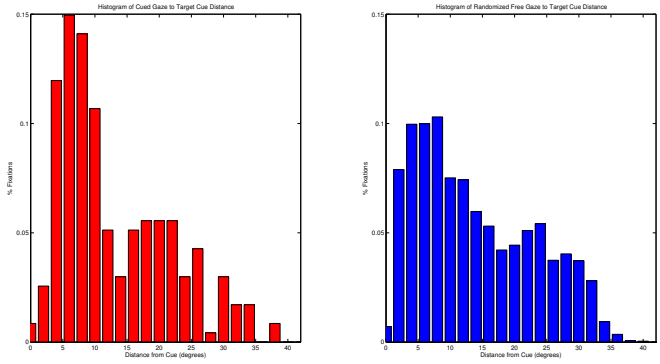
Histogram distribution of Esperanza 1 Image



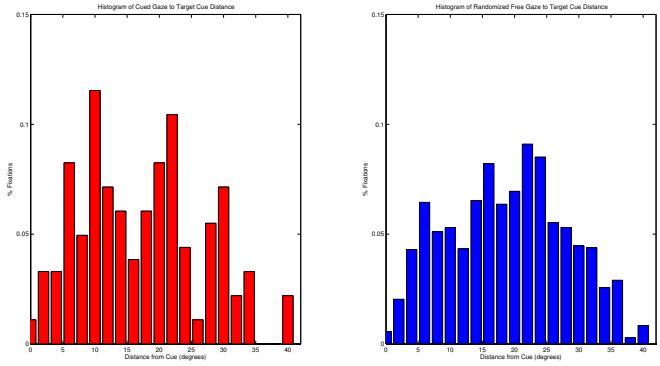
Histogram distribution of Esperanza 2 Image



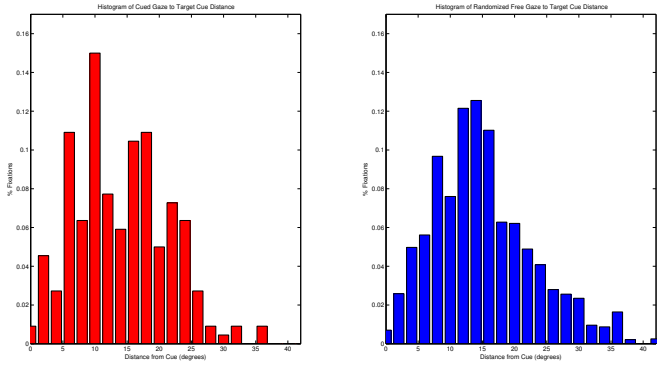
Histogram distribution of Esperanza 3 Image



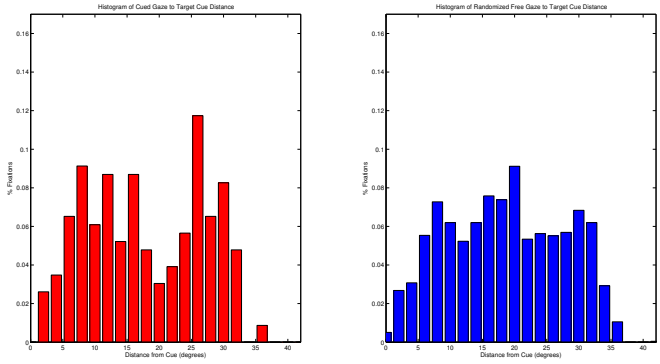
Histogram distribution of Esperanza 4 Image



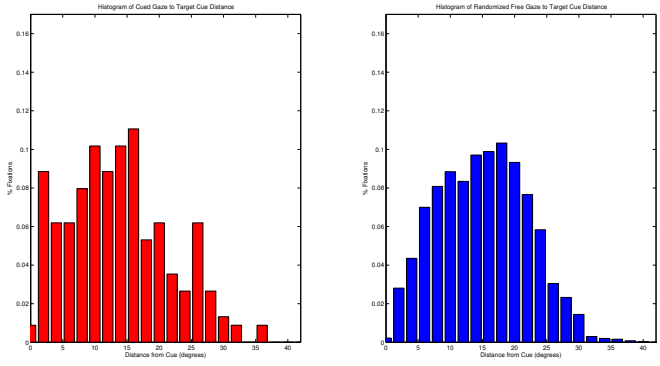
Histogram distribution of Rondonia 1 Image



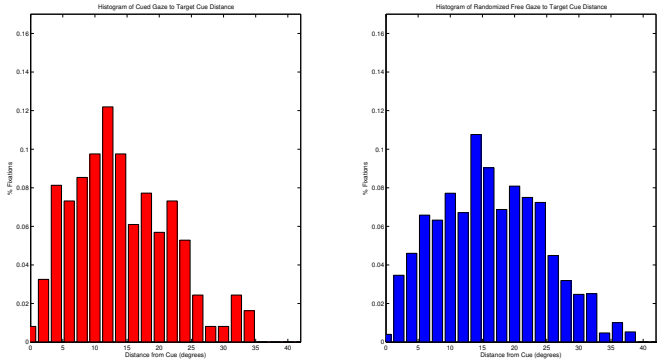
Histogram distribution of Rondonia 2 Image



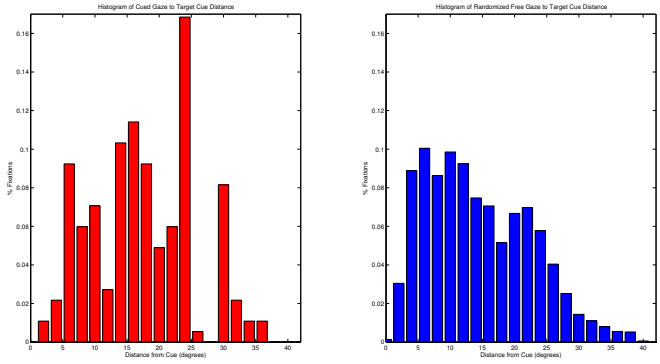
Histogram distribution of Rondonia 3 Image



Histogram distribution of Rondonia 4 Image

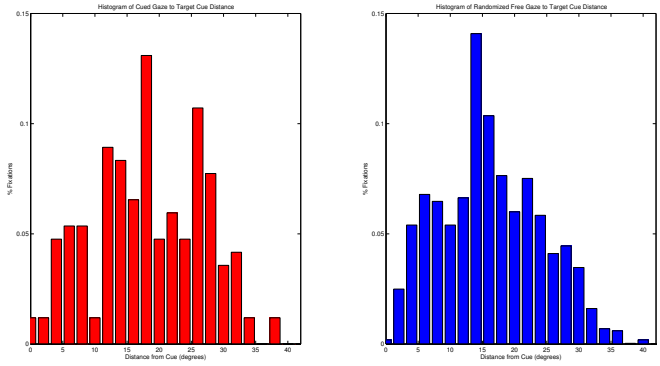


Histogram distribution of Salvador 1 Image

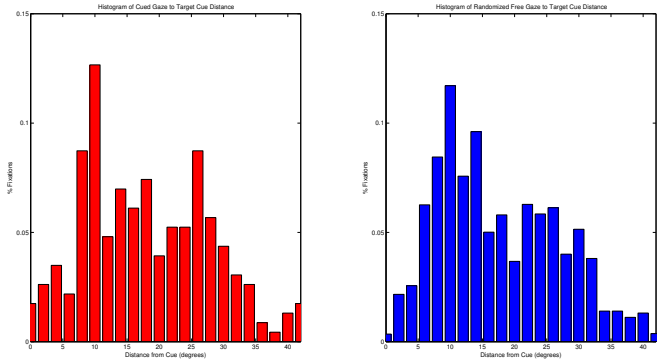


Histogram distribution of Salvador 2 Image

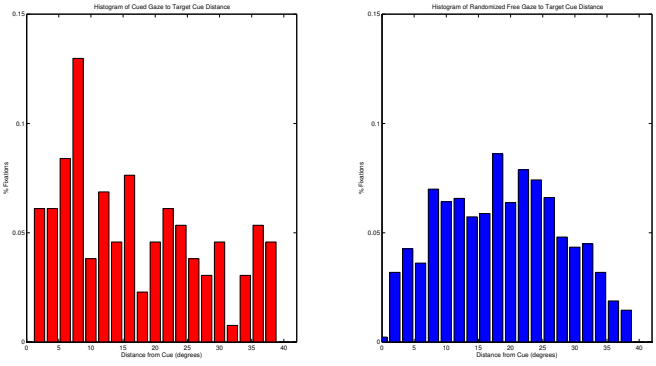
Histograms for Cue Durations of 3501-5000ms



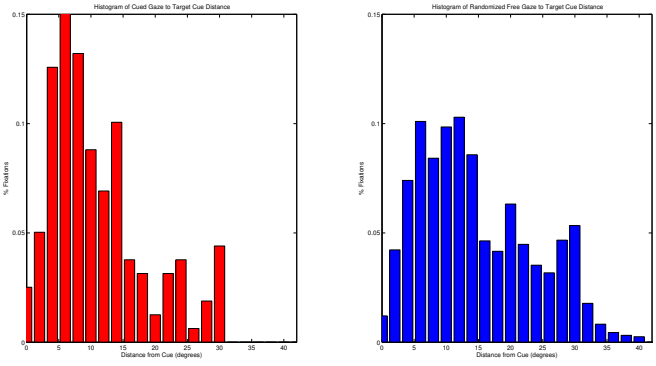
Histogram distribution of Esperanza 1 Image



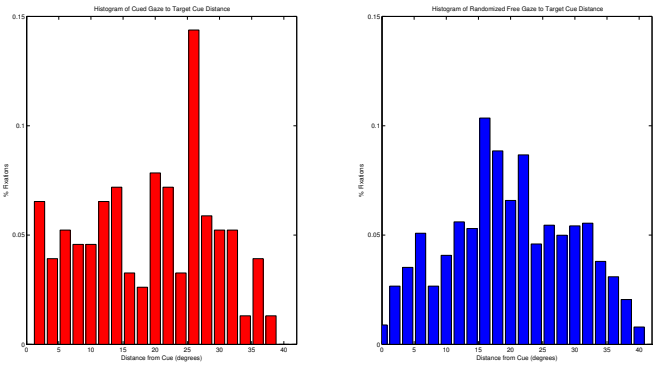
Histogram distribution of Esperanza 2 Image



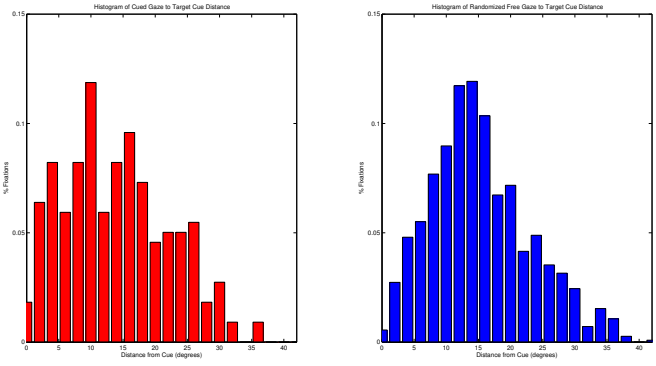
Histogram distribution of Esperanza 3 Image



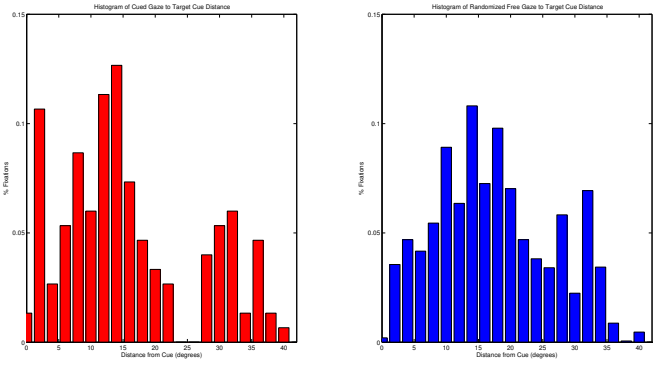
Histogram distribution of Esperanza 4 Image



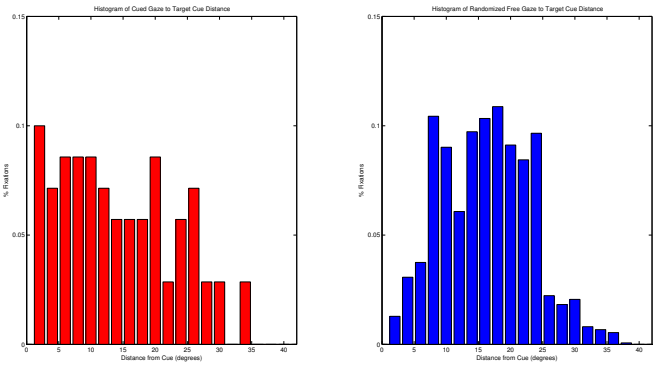
Histogram distribution of Rondonia 1 Image



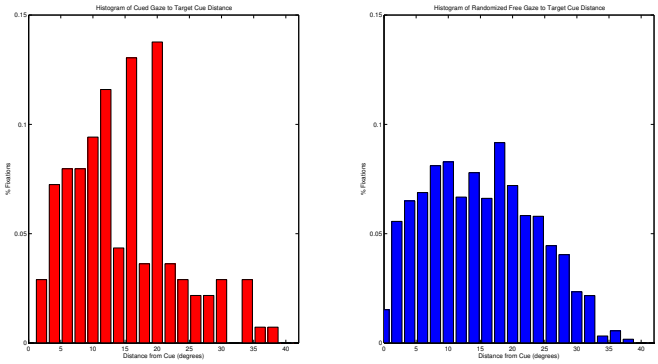
Histogram distribution of Rondonia 2 Image



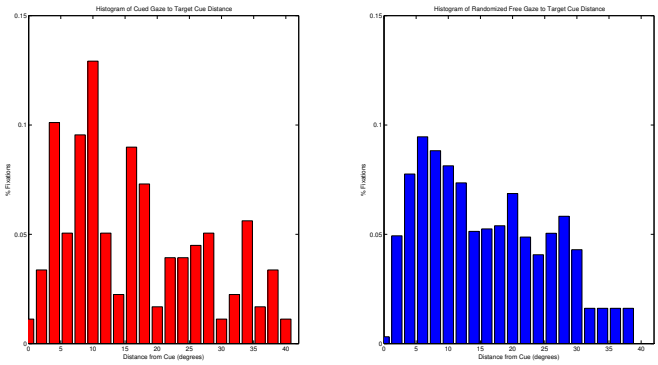
Histogram distribution of Rondonia 3 Image



Histogram distribution of Rondonia 4 Image



Histogram distribution of Salvador 1 Image



Histogram distribution of Salvador 2 Image